

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 08:55:18 ; Search time 17.52 Seconds  
(without alignments)  
1590.520 Million cell updates/sec

Title: US-10-041-006-7

Perfect score: 1371

Sequence: 1 MRPAVPLLLLLCFCGSGRA.....WIRRIPLKQFQPARLGCK 290

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	618.5	39.4	343	1 A57014	proctasin (EC 3.4.1.1)
2	594	37.8	270	2 S6160	mast cell tryptase (EC 3.4.21.1)
3	588.5	37.5	275	2 B35863	tryptase (EC 3.4.21.1)
4	587.5	37.4	275	2 A35863	tryptase (EC 3.4.21.1)
5	584.5	37.2	275	2 C35863	tryptase (EC 3.4.21.1)
6	581	37.0	274	2 A45754	tryptase (EC 3.4.21.1)
7	578.5	36.8	273	2 A47246	tryptase (EC 3.4.21.1)
8	565.5	36.0	274	2 J04171	mast cell proteinase (EC 3.4.21.1)
9	564	35.9	276	2 A36654	tryptase (EC 3.4.21.1)
10	560.5	35.7	275	2 A32410	plasma kallikrein (EC 3.4.21.1)
11	533	33.9	638	1 KOHUP	plasma kallikrein (EC 3.4.21.1)
12	526	33.5	810	2 B30848	plasma kallikrein (EC 3.4.21.1)
13	518	33.0	638	1 KOMSPL	hepsin (EC 3.4.21.1)
14	517.5	32.9	417	1 S00845	plasma kallikrein (EC 3.4.21.1)
15	517	32.9	638	1 KORPPL	plasma kallikrein (EC 3.4.21.1)
16	513.5	32.7	237	2 S6702	tryptase (EC 3.4.21.1)
17	509	32.4	455	2 A61545	hepsin (EC 3.4.21.1)
18	505.5	32.2	416	1 PLHV	plasma (EC 3.4.21.1)
19	502	32.0	810	1 PLHO	plasma (EC 3.4.21.1)
20	496.5	31.6	812	1 PLBO	mastocytoma proteinase (EC 3.4.21.1)
21	493	31.4	269	2 B32410	enteropeptidase (EC 3.4.21.1)
22	491.5	31.3	1034	1 A53663	coagulation factor (EC 3.4.21.1)
23	487.5	31.0	625	1 KPHOI	plasma (EC 3.4.21.1)
24	479.5	30.5	812	1 PLMS	plasma (EC 3.4.21.1)
25	478	30.4	790	1 PLPG	plasma (EC 3.4.21.1)
26	475	30.4	460	2 B61545	apolipoprotein (a) (EC 3.4.21.1)
27	477.5	30.4	1420	2 A32869	apolipoprotein (a) (EC 3.4.21.1)
28	469	29.9	1524	2 T30337	apolipoprotein (a) (EC 3.4.21.1)
29	466	29.7	1524	2 S00657	apolipoprotein (a) (EC 3.4.21.1)

## ALIGNMENTS

30	462.5	29.4	855	2 JC7731	membrane-bound arg
31	462.5	29.4	855	2 JC7735	membrane type-seri
32	460.5	29.3	1113	2 JC7315	low-density lipopr
33	459.5	29.2	1035	1 A43090	enteropeptidase (E
34	457	29.1	1019	1 A56318	enteropeptidase (E
35	454	28.9	810	2 I46260	plasma (EC 3.4.21
36	448	28.5	415	1 A34170	acrosin (EC 3.4.21
37	444.5	28.3	263	1 KYRTB	chymotrypsin (EC 3
38	442.5	28.2	1004	2 T30338	oviductin (EC 3.4
39	441.5	28.1	263	2 S47537	chymotrypsin (EC 3
40	441.5	28.1	367	2 JE0104	testicular serine
41	440.5	28.0	263	2 A21195	chymotrypsin (EC 3
42	438	27.9	431	2 S47538	acrosin (EC 3.4.21
43	435.5	27.7	436	2 JX0172	acrosin (EC 3.4.21
44	433	27.6	786	1 A47547	serine proteinase
45	432.5	27.5	275	2 S40005	tryptase (EC 3.4.21

## RESULT 1

A57014  
proctasin (EC 3.4.21.1) precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 18-Jun-1999  
C:Accession: A57014; A54866  
R:Yu, J.X.; Chao, L.; Chao, J.  
J. Biol. Chem. 270, 13483-13489, 1995  
A:Title: Molecular cloning, tissue-specific expression, and cellular localization of  
A:Reference number: A57014; M01D:95286644  
A:Accession: A57014  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-343 <RPS>  
A:Cross-references: GB:L4151; NID:9862304; PIDN:AAC41759.1; PID:9862305  
A:Experimental source: prostate  
A:Note: parts of this sequence were determined by protein sequencing  
R:Yu, J.X.; Chao, L.; Chao, J.  
J. Biol. Chem. 269, 18843-18848, 1994  
A:Title: Proctasin is a novel human serine proteinase from seminal fluid. Purification  
A:Reference number: A54866; M01D:94308140  
A:Accession: A54866  
A:Molecule type: protein  
A:Residues: 45-64 <YDA>  
A:Genetics:  
A:Gene: GDB:PRSS8  
A:Cross-references: GDB:676446; OMIM:600823  
A:Map position: 16p11.2-16p11.2  
C:Superfamily: proctasin; trypsin homology  
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein  
F:1-32/Domain: signal sequence #status predicted <SIG>  
F:33-44/Domain: proctasin heavy chain #status predicted <CHL>  
F:33-44/Domain: proctasin light chain #status predicted <CHL>  
F:45-343/Domain: trypsin homology <TRY>  
F:45-281/Domain: trypsin homology #status predicted <TM1>  
F:323-344/Domain: transmembrane #status predicted <TM2>  
F:37-154/70-86,168-244,201-223,234-262/Disulfide bonds: #status predicted  
F:85,134,238/Active site: His, Asp, Ser #status predicted  
F:159/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 39.4% Score 618.5; DB 1; Length 343;  
Best Local Similarity 42.9% Pred. No. 6.5e-47;  
Matches 124; Conservative 47; Mismatches 99; Indels 19; Gaps 7;

DB 6 AVPLLL--LTCFGSQRKAAATACG--RPPMLRNWVGODTQEGEMPMQVSIORNGSHRCG 61  
14 AVAILLYLCGLSSGABEAPCGVAPCA--RITGGSSAVAGCPWQVSTIREGVHCG 71  
OY 62 GSILAEQWVLTAAHCFRNTSETSLYQVILGARQVQPGPHANTARVQVESPLLYQGTAS 121  
DB 72 GSILVSGWVLSAHCAPPSSEHKEAYEVKLGIAHQDLSYSEDAKVSTLKDIIPHPYLBGS 131

**RESULT 3**  
 B35863  
 tryptase (EC 3.4.21.59) II precursor - human  
 N:Alternate names: tryptase beta  
 C:Species: Homo sapiens (man)  
 C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 21-Jul-2000  
 C:Accession: B35863; A37193; I56473  
 R:VanderSlice, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990  
 A:Title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine pr  
 A:Reference number: A35863; MUID:90251647  
 A:Accession: B35863  
 A:Molecule type: mRNA; DNA  
 A:Residues: 1-275 <YAN>  
 A:Cross-references: GB:NM33492; NID:g339982; PIDN:AAA36779.1; PID:g339983

tryptase (EC 3.4.21.59) I precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 12-Oct-1990 #sequence revision 12-Oct-1990 #text\_change 21-Jul-2000  
C:Accession: A35863; D35863; A60939; A39326  
R:Vanderslice, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey,  
Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990  
A:Title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine  
A:Reference number: A35863, MUID:90251647  
A:Accession: A35863  
A:Molecule type: DNA  
A:Residues: 1-275 <VAN>  
A:Cross-references: GB:A33494; NID:g3327804; PIDN:AAC83172.1; PID:g333977  
A:Accession: D35863  
A:Molecule type: mRNA



Query Match 37.0%; Score 581; DB 2; Length 274;  
 Best Local Similarity 43.8%; Pred. No. 1e-43;  
 Matches 124; Conservative 38; Mismatches 93; Indels 28; Gaps 5;

OY 9 LLLLLCFSQAKAKATACGRPRMLNRMYGGODTOEGEMPQVSIQ---RNGSHFGGSLI 65  
 DB 5 LLLALLPVLASAYAPAPVQALQAGIVGGEAPRSKMPQVSLRVDRTVMHFCGSSLI 64  
 OY 66 AEQWVLTAAHCFRNTSETSLYQVLLG-----ARQLVQCPHANY---ARQVQESNPL 115  
 DB 65 HPQWVLTAAHCFRNTSETSLYQVLLG-----LGDVDVLDLALRYNSGTHLYDQDLLPVSRIYHPQ 111  
 OY 116 YQGAASADVALVELAPVFTNYILPVCPLDPSVIFETGMNCWVGSGSPSEDLLEP 175  
 DB 112 FYIITQADIALLELEPVSNSRVHTVMLPRASEFPFPMPCWVGMDVNDLEPLRP 171  
 OY 176 RIILQKLVPIIDPRKMLLSKQTEFGYQKTIKNDMLCAGFEKGKDKAGSGGGLVC 235  
 DB 172 FPLQVAVPPIEMENHTCAKTHLGAVTGDVRIIRIDMLCAG--NSQDSCGSGGGLVC 229  
 OY 236 LVGOSMLQAGVISMGEGCARQNRPGVYIRYTAHNNHRIIRPK 278  
 DB 230 KVNCTWLAGVSMDEGCAPNRPGITRTYTYLDWTHHIVPK 272

RESULT 7  
 A47246  
 tryptase (EC 3.4.21.59) 2 - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000  
 C:Accession: A47246  
 R:McNell, H.P.; Reynolds, D.S.; Schiller, V.; Ghildyal, N.; Gurely, D.S.; Austen, K.F.;  
 Proc. Natl. Acad. Sci. U.S.A. 89, 11174-11178, 1992  
 A:Title: Isolation, characterization, and transcription of the gene encoding mouse mast  
 A:Reference number: A47246; MUID:93087489  
 A:Accession: A47246  
 A:Status: preliminary  
 A:Molecule type: nucleic acid  
 A:Residues: 1-273 <MCN>  
 A:Cross-references: GB:L00653; NID:9200518; PID:AAA39982.1; PID:9200519  
 A:Note: Sequence extracted from NCBI backbone (NCBI:119745, NCBI:119746)  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: hydrolase; serine proteinase  
 F:29-265/Domain: trypsin homology <TRY>

Query Match 36.8%; Score 578.5; DB 2; Length 273;  
 Best Local Similarity 45.5%; Pred. No. 1.7e-43;  
 Matches 125; Conservative 41; Mismatches 96; Indels 13; Gaps 7;

OY 9 LLLLLCFSQAKAKATACGRPRMLNRMYGGODTOEGEMPQVSIQ---RNGSGSL 64  
 DB 5 LLLTLPPLSLVHAAPG---PAMTREGIVGGEAHGNKMPQVSLRANDTVMHFCGSSLI 61  
 OY 65 IAEQWVLTAAHCFRNTSETSLYQVLLGARQLVQCPHANYRQVESNPLVQGTASSA 123  
 DB 62 IHPQWVLTAAHCFRNTSETSLYQVLLGARQLVQCPHANYRQVESNPLVQGTASSA 118  
 OY 124 DVALVELAPVFTNYILPVCPLDPSVIFETGMNCWVGSGSPSEDLLEPRLIQLAV 183  
 DB 119 DIALKLTNPVNIISYVHVPPLPRASEFPFPMPCWVGMDVNDLEPLRP 178  
 OY 184 PIIDPRKMLLSKQTEFGYQKTIKNDMLCAGFEKGKDKAGSGGGLVC 243  
 DB 179 PIIEHLDLAKTKHKLITGDVNIHYRDMILCAG--HDSQGGSGGLVCXVEDTWLQ 236  
 OY 244 AGVISMGEGCARQNRPGVYIRYTAHNNHRIIRPK 278  
 DB 237 AGVSMGEGCARQNRPGITRTYTYLDWTHHIVPK 271

RESULT 8  
 JC4171

tryptase (EC 3.4.21.59) precursor - rat  
 N:Alternate names: mast cell tryptase  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 27-Aug-1995 #sequence\_revision 27-Oct-1995 #text\_change 20-Jun-2000  
 C:Accession: JC4171  
 R:Ido, H.; Itoh, H.; Tomita, M.; Murakumo, Y.; Kobayashi, T.; Maruyama, H.; Osada, Y.  
 J. Biochem. 118, 210-215, 1995  
 A:Title: cDNA sequencing and expression of rat mast cell tryptase.  
 A:Reference number: JC4171; MUID:96015171  
 A:Accession: JC4171  
 A:Molecule type: mRNA  
 A:Residues: 1-274 <IDE>  
 A:Cross-references: DDBJ:D38455; NID:9556555; PIDN:BA07486.1; PID:9556556  
 C:Comment: This enzyme is basically specific for a connective tissue mast cell, it is  
 elase inhibitors.  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: glycoprotein; hydrolase; mast cell; serine proteinase; zymogen  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-29/Domain: activation peptide #status predicted <ACT>  
 F:30-274/Product: mast cell tryptase #status predicted <MAT>  
 F:30-266/Domain: mast cell tryptase homology <TRY>  
 F:73,120,223/Active site: His, Asp, Ser #status predicted  
 F:131/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 36.0%; Score 565.5; DB 2; Length 274;  
 Best Local Similarity 43.7%; Pred. No. 2.4e-42;  
 Matches 122; Conservative 41; Mismatches 97; Indels 19; Gaps 7;

OY 9 LLLLLCFSQAKAKATACGRPRMLNRMYGGODTOEGEMPQVSIQ---RNGSGSLI 65  
 DB 4 LLLTLPPLSLVHAAPG---PAMTREGIVGGEAHGNKMPQVSLRANDTVMHFCGSSLI 63  
 OY 66 AEQWVLTAAHCFRNTSETSLYQVLLGARQLVQCPHANYRQVESNPLVQGT 120  
 DB 64 HPQWVLTAAHCFRNTSETSLYQVLLGARQLVQCPHANYRQVESNPLVQGT 116  
 OY 121 SSADVALVELAPVFTNYILPVCPLDPSVIFETGMNCWVGSGSPSEDLLEPRLIQL 179  
 DB 117 DGADIALLELEPVSNSRVHTVMLPRASEFPFPMPCWVGMDVNDLEPLRP 175  
 OY 180 KLVAVPIIDPRKMLLSKQTEFGYQKTIKNDMLCAGFEKGKDKAGSGGGLVC 239  
 DB 176 QKVPVIVENSICDRKRYHTGTYGDVPIYQDMLCAG--NTRSDSCGSGGGLVCVK 233  
 OY 240 SMLQAGVISMGEGCARQNRPGVYIRYTAHNNHRIIRPK 278  
 DB 234 TWLQAGVSMGEGCARQNRPGITRTYTYLDWTHHIVPK 272

RESULT 9  
 A38654  
 mast cell proteinase 6 (EC 3.4.21.-) precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 21-Feb-1992 #sequence\_revision 17-Feb-1994 #text\_change 22-Jun-1999  
 C:Accession: A38654; B38654; D35646; I59478  
 R:Reynolds, D.S.; Gurely, D.S.; Austen, K.F.; Serafin, W.E.  
 J. Biol. Chem. 266, 3847-3853, 1991  
 A:Title: Cloning of the cDNA and gene of mouse mast cell protease-6. Transcription by  
 A:Reference number: A38654; MUID:91139682  
 A:Accession: A38654  
 A:Molecule type: DNA  
 A:Residues: 1-276 <REY>  
 A:Cross-references: GB:M57625; NID:9200506; PIDN:AAA39987.1; PID:9200507  
 A:Note: the authors translated the codon CGC for residue 24 as Ala, GAG for residue 3  
 s Gly, GAG for residue 148 as Gly, GAG for residue 168 as Gly, and GAA for 185 as Gly  
 A:Accession: B38654  
 A:Molecule type: mRNA  
 A:Residues: 1-276 <REY>  
 A:Cross-references: NID:9200508; PIDN:AAA39988.1; PID:9200509  
 R:Reynolds, D.S.; Stevens, R.L.; Lane, W.S.; Carr, M.H.; Austen, K.F.; Serafin, W.E.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 3230-3234, 1990  
 A:Title: Different mouse mast cell populations express various combinations of at least





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 08:52:53 ; Search time 13.11 Seconds  
(without alignments)  
540,307 Million cell updates/sec

Title: US-10-041-006-7  
Perfect score: 1571  
Sequence: 1 MRRAVAVPRLLLLCFSGSQA.....WIRRIIPKIQFQPARLGQK 290

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 231628  
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PTCUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Score	Query Match	Length	ID	Description
1	1197	76.2	304	4	US-09-088-651-2 Sequence 2, Appl
2	602.5	38.4	299	4	US-08-944-483-66 Sequence 66, Appl
3	594	37.8	270	2	US-08-978-404B-8 Sequence 8, Appl
4	590.5	37.6	273	2	US-08-978-404B-6 Sequence 6, Appl
5	588.5	37.5	274	2	US-09-016-366A-21 Sequence 21, Appl
6	588.5	37.5	274	2	US-08-978-404B-16 Sequence 16, Appl
7	587.5	37.4	273	2	US-09-016-366A-19 Sequence 19, Appl
8	587.5	37.4	273	2	US-08-978-404B-14 Sequence 14, Appl
9	578.5	36.8	267	2	US-09-016-366A-23 Sequence 23, Appl
10	578.5	36.8	267	2	US-08-978-404B-18 Sequence 18, Appl
11	578.5	36.8	273	2	US-08-978-404B-3 Sequence 3, Appl
12	574	36.5	314	4	US-09-008-271A-3 Sequence 3, Appl
13	573.5	36.5	249	4	US-09-079-970A-5 Sequence 5, Appl
14	572.5	36.4	275	2	US-09-016-366A-17 Sequence 17, Appl
15	572.5	36.4	275	2	US-08-978-404B-12 Sequence 12, Appl
16	568.5	36.2	245	4	US-09-079-970A-6 Sequence 6, Appl
17	565.5	36.0	274	2	US-08-978-404B-5 Sequence 5, Appl
18	564	35.9	276	2	US-09-016-366A-15 Sequence 15, Appl
19	564	35.9	276	2	US-08-978-404B-21 Sequence 21, Appl
20	560.5	35.7	245	4	US-08-944-483-69 Sequence 69, Appl
21	527.5	33.6	248	4	US-08-944-483-63 Sequence 63, Appl
22	523.5	33.3	416	2	US-09-000-846-2 Sequence 2, Appl
23	517	32.9	638	2	US-08-681-151-3 Sequence 3, Appl
24	505	32.1	791	1	US-08-643-219-1 Sequence 1, Appl
25	505	32.1	791	3	US-08-851-350-1 Sequence 1, Appl
26	504	32.1	256	2	US-09-027-337-3 Sequence 3, Appl
27	503	32.0	814	1	US-08-750-711-1 Sequence 1, Appl

28	502	32.0	790	1	US-08-469-486-54 Sequence 54, Appl
29	502	32.0	790	2	US-08-469-658-54 Sequence 54, Appl
30	502	32.0	791	2	US-09-131-995-1 Sequence 1, Appl
31	502	32.0	791	2	US-08-832-087B-1 Sequence 1, Appl
32	502	32.0	791	4	US-09-132-134-1 Sequence 1, Appl
33	502	32.0	810	1	US-07-854-603-2 Sequence 2, Appl
34	502	32.0	810	1	US-08-147-000B-29 Sequence 29, Appl
35	502	32.0	810	1	US-08-086-514-1 Sequence 1, Appl
36	502	32.0	810	6	US-09-0340-8 Patent No. 5200340
37	499	31.8	255	4	US-08-944-483-67 Sequence 67, Appl
38	499	31.8	546	6	US-09-040-6 Patent No. 5200340
39	493	31.4	269	2	US-08-978-404B-10 Sequence 10, Appl
40	483.5	30.8	230	1	US-08-456-840-47 Sequence 47, Appl
41	483.5	30.8	230	1	US-08-266-407A-47 Sequence 47, Appl
42	483.5	30.8	230	2	US-08-892-544-47 Sequence 47, Appl
43	483.5	30.8	230	2	US-08-766-983-12 Sequence 12, Appl
44	483.5	30.8	230	4	US-08-944-483-53 Sequence 53, Appl
45	483.5	30.8	230	4	US-09-296-219-12 Sequence 12, Appl

## ALIGNMENTS

RESULT 1  
US-09-088-651-2  
; Sequence 2, Application US/09088651  
; Patent No. 6165771  
GENERAL INFORMATION:  
APPLICANT: BURGESS, NICOLA A.  
APPLICANT: CLINKENBERG, HELEN E.  
APPLICANT: SOUTHAN, CHRISTOPHER D.  
TITLE OF INVENTION: NOVEL COMPOUNDS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: RAINIER & PRESTIA  
STREET: P.O. BOX 980  
CITY: VALLEY FORGE  
STATE: PA  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/088,651  
FILING DATE: JUNE 1, 1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB9712088.5  
FILING DATE: 10-JUNE-1997  
APPLICATION NUMBER: EP 97308295.1  
FILING DATE: 17-OCT-1997  
APPLICATION NUMBER: GB 9803650.2  
FILING DATE: 20-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F.  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GH30358  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0701  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 304 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-088-651-2

Query Match 76.2%; Score 1197; DB 4; Length 304;  
 Best Local Similarity 82.1%; Pred. No. 1.5e-124;  
 Matches 234; Conservative 7; Mismatches 36; Indels 8; Gaps 3;

QY 1 MRRPAVPLLLLCGSGORAKATACGRPMILNRVYGODTQEGEMPMQVSIORNGSHFC 60  
 DB 1 MRRPAVPLLLLCGSGORAKATACGRPMILNRVYGODTQEGEMPMQVSIORNGSHFC 60  
 QY 61 GGSLLAEQWVLTAAHCENFTSETSLYQVLLARQLVQGPPIAMTARVQVESNPLYOGTA 120  
 DB 61 GGSLLAEQWVLTAAHCENFTSETSLYQVLLARQLVQGPPIAMTARVQVESNPLYOGTA 120  
 QY 121 SSADVALVELAPVFTYIIPVCLDPDSVIFETGMNCWVGWSPSEEDLLPEPRILQK 180  
 DB 121 SSADVALVELAPVFTYIIPVCLDPDSVIFETGMNCWVGWSPSEEDLLPEPRILQK 180  
 QY 181 LAVPIIDTPKCNLYSKDTERGYQPKTIKNDMLCAGFEKCKDCKGSDGPLYCLVGOS 240  
 DB 181 LAVPIIDTPKCNLYSKDTERGYQPKTIKNDMLCAGFEKCKDCKGSDGPLYCLVGOS 240  
 241 WLQAGV-ISMGE-----CARQNPQVYIRTAHNNHRIIPKL 279  
 239 FQPIGPGGKGLGPPVWMLQTEPPSVILRVNGOQNMTHRIIPKL 283

RESULT 2  
 US-08-944-483-66  
 Sequence 66, Application US/08944483  
 Patent No. 6232456

GENERAL INFORMATION:  
 APPLICANT: COHEN, MAURICE  
 APPLICANT: COLPITTS, TRACEY L.  
 APPLICANT: FRIEDMAN, PAULA N.  
 APPLICANT: GRANADOS, EDWARD N.  
 APPLICANT: KLAS, MICHAEL R.  
 APPLICANT: RUSSELL, JOHN C.  
 APPLICANT: STEWART, KENT D.  
 APPLICANT: STROUPE, STEVEN D.  
 TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS  
 TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
 NUMBER OF SEQUENCES: 76  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Abbott Laboratories  
 STREET: 100 Abbott Park Road  
 CITY: Abbott Park  
 STATE: IL  
 COUNTRY: USA  
 ZIP: 60064-3500

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/944,483  
 FILING DATE:  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:  
 NAME: Becker, Cheryl L.  
 REGISTRATION NUMBER: 35,441  
 REFERENCE/DOCKET NUMBER: 6183.US.01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 847/935-1729  
 TELEFAX: 847/938-2623

TELEX:  
 INFORMATION FOR SEQ ID NO: 66:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 299 amino acids

TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: No. 6232456e  
 US-08-944-483-66

Query Match 38.4%; Score 602.5; DB 4; Length 299;  
 Best Local Similarity 43.8%; Pred. No. 1.2e-58;  
 Matches 112; Conservative 44; Mismatches 87; Indels 13; Gaps 4;

QY 35 MGGQDQEGEMPMQVSIORNGSHFCGSLIAEOWLTAAHCENFTSETSLYQVLLARQ 94  
 DB 1 ITGGSSAVAGQWQVSTIEGYHVCGLSVQWVLSAHCENFTSETSLYQVLLARQ 60  
 QY 95 LVQGPPIAMTARVQVESNPLYOGTASADVALVELAPVFTYIIPVCLDPDSVIFET 154  
 DB 61 LQSVSEDAKSTLKDILPHPSYLOEGSGDIALQLSRPTFSRYIRPCLPAANASFPN 120  
 QY 155 GNMCCWVGWSPSEEDLLPEPRILQKLAVIDPDKCNLYSKDTERGYQPKTIKNDM 212  
 DB 121 GLHCTVYGMGHVAPSVSLTPKP--LQQLVPLISRETQNCILNIDAK--PEEPHFVQEDM 177  
 QY 213 LCAGFEKCKDCKGSDGPLYCLVGOSWLVAGVISGEGCARQNPQVYIRTAHNNHRI 272  
 DB 178 VCAGYVEGKDKACQDGSGLPSCVPEGILYLGIVSGMDACARRNPVYITLASSYASNI 237  
 QY 273 H-----RIIPKLQ 280  
 DB 238 QSKVTELQPRVVPQTQ 253

RESULT 3  
 US-08-978-404B-8  
 Sequence 8, Application US/08978404B  
 Patent No. 5968782

GENERAL INFORMATION:  
 APPLICANT: Stevens, Richard L.  
 TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
 TITLE OF INVENTION: FIBRINOGEN  
 NUMBER OF SEQUENCES: 74  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
 STREET: 600 Atlantic Avenue  
 CITY: Boston  
 STATE: MA  
 COUNTRY: U.S.A.  
 ZIP: 02210-2211

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/978,404B  
 FILING DATE: 25-NOV-97  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/032,354  
 FILING DATE: 04-DEC-1996

ATTORNEY/AGENT INFORMATION:  
 NAME: Plummer, Elizabeth R.  
 REGISTRATION NUMBER: 36,637  
 REFERENCE/DOCKET NUMBER: B0801/7090  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-720-3500  
 TELEFAX: 617-720-2441

TELEX:  
 INFORMATION FOR SEQ ID NO: 8:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 270 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single



TOPOLOGY: linear  
MOLECULE TYPE: No. 5968782e  
US-08-978-404B-8

Query Match 37.8%; Score 594; DB 2; Length 270;  
Best Local Similarity 44.0%; Pred. No. 9.1e-58;  
Matches 122; Conservative 48; Mismatches 87; Indels 20; Gaps 7;

QY 6 AVPLLLLLCGESORAKATACGRRLNRMVGGODTOEGEMPWQVSIOHNGS---HFCGG 62  
DB 8 ALPLFSLM-----HRSPLCOENG-----IVGQEARGNKMPQVSLRANETWMHFCGG 56  
QY 63 SLIAEOWVLTAAHCFRNT-SETSLYOVLLGAROLVOPGHAMTARYKQVESNPLYOGTAS 121  
DB 57 SLIHQWVLTAAHCFVPTIADPNKRVOLRKQYLYY---HDHLLAVSRITHTPEYATON 113  
122 SADVALVLEAPVFTYIIPVCLDPDSVTFEFGMNCWYTGWGSPEEDLLPEPRILQKL 181  
114 GADIALLELKNPVNISHVHPVSLPPASETFPSGTLCTWYTGWGNINDVSLPPFPPLKEV 173  
QY 182 AVPIIDFKCNILYSKDTFEGYOPRTIKNDMLCAGFEGRKADKDGSGPLVCLVQGSW 241  
DB 174 QVPVVENLCLDKHKYKGYTDNTHIVRDMCLAG-NEG-HDSCQDGSGLVCKVNGTW 231  
QY 242 IQAGVTSWEGECARONRGVYIRTAHNMTHRIIPK 278  
DB 232 IQAGVTSWEGECALPNRPGIYTRVYLYLDMITHRYVPK 268

## RESULT 4

US-08-978-404B-6  
Sequence 6, Application US/08978404B

Patent No. 5968782

GENERAL INFORMATION:

APPLICANT: Stevens, Richard L.

TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES

TITLE OF INVENTION: FIBRINOGEN

NUMBER OF SEQUENCES: 74

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wolf, Greenfield & Sacks, P.C.

STREET: 600 Atlantic Avenue

CITY: Boston

STATE: MA

COUNTRY: U.S.A.

ZIP: 02210-2211

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/978,404B

FILING DATE: 25-NOV-97

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/032,354

FILING DATE: 04-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Plumer, Elizabeth R.

REGISTRATION NUMBER: 36,637

REFERENCE/DOCKET NUMBER: B0801/7090

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441

TELEX:

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 273 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 5968782e

US-08-978-404B-6

Query Match 37.6%; Score 590.5; DB 2; Length 273;  
Best Local Similarity 45.6%; Pred. No. 2.3e-57;  
Matches 125; Conservative 43; Mismatches 95; Indels 11; Gaps 6;

QY 9 LLLLCGSGRAKATACGRRLNRMVGGODTOEGEMPWQVSIOHNGS---HFCGSLI 65  
DB 5 LLLTPLLSTLVHAAPSLAMP--EGVGGQEARGNKMPQVSLRANETWMHFCGSLI 62  
QY 66 AEQWVLTAAHCF-RNTSETSLYOVLLGAROLVOPGHAMTARYKQVESNPLYOGTAS 124  
DB 63 HPQWVLTAAHCFVPTIADPNKRVOLRKQYLYY---HDHLLAVSRITHTPEYATON 119  
QY 125 VALVLEAPVFTYIIPVCLDPDSVTFEFGMNCWYTGWGSPEEDLLPEPRILQKL 184  
DB 120 IALLKLNPNVITSNVHTVSLPPASETFPSGTLCTWYTGWGNINDVSLPPFPPLKEV 179  
QY 185 IIDPFKCNILYSKDTFEGYOPRTIKNDMLCAGFEGRKADKDGSGPLVCLVQGSW 244  
DB 180 IVENRLCLDKHKYKGYTDNTHIVRDMCLAG-NEG-HDSCQDGSGLVCKVNGTW 237  
QY 245 QVTSWEGECARONRGVYIRTAHNMTHRIIPK 278  
DB 238 QVTSWEGECALPNRPGIYTRVYLYLDMITHRYVPK 271

## RESULT 5

US-09-016-366A-21  
Sequence 21, Application US/09016366A

Patent No. 5955431

GENERAL INFORMATION:

APPLICANT: Stevens, Richard L.

TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE

TITLE OF INVENTION: INHIBITORS

NUMBER OF SEQUENCES: 65

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wolf, Greenfield & Sacks, P.C.

STREET: 600 Atlantic Avenue

CITY: Boston

STATE: MA

COUNTRY: U.S.A.

ZIP: 02210-2211

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/016,366A

FILING DATE: January 30, 1998

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/037,090

FILING DATE: 05-FEB-1997

ATTORNEY/AGENT INFORMATION:

NAME: Plumer, Elizabeth R.

REGISTRATION NUMBER: 36,637

REFERENCE/DOCKET NUMBER: B0801/7093

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441

TELEX:

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:

LENGTH: 274 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-016-366A-21

Best Local Similarity 43.8%; Pred. No. 3.8e-57;  
Matches 120; Conservative 46; Mismatches 99; Indels 9; Gaps 4;

4 LLALPVLASRAYAAPRGALQVRVIGGQENRDKRPQVSLRVKHGPRVMHFCGGSLI 63

66 AEQWLTAAHCF-RNMSLTYVLLGARGLVQGPHAMYARVRQVESNPPLYQGTASSAD 124

125 VALVELAEVPRTNYILPYCLBPSVIFPTGNCWYTGWGSSEPNILBPPPOKTYND 184

[illegible][illegible]

101 LMEHHCLDNIHSGAIGDDPKIKVDMLCAG--NRDSCQGGSGGPLVCKVNGTWLQA 238

245 GVVISWGGECARONRBPQVYIRVTAHHNHTPIPTDY 278

239 GVSVWGEGCAQPNRPGIYTRVYYLIDWIIHHYVPK 212

2  
2011

Sequence 19, Application NS/09016366A  
-09-016-366A-19

Patent No. 5955431

APPLICANT: Stevens, Richard L.  
APPLICANT: Huang, Chifu  
TITLE OF INVENTION:

TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
NUMBER OF SEQUENCES: 65  
INHIBITORS

ADDRESSEE: Wolf, Greenfield & Sacks, P.C.

STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA

STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02310-2211

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
COMMUNICATIONS: 1200 BPS

SOFTWARE FASTSEQ FOR Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016 366A

INFORMATION CONTAINED HEREIN IS UNCLASSIFIED  
 DATE 03/05/01 BY 60324 UCBAW/STP  
 FILING DATE: January 30, 1998  
 CLASSIFICATION: 530

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/037,090  
FILING DATE:

NAME: Plummer, Elizabeth D  
FILING DATE: 05-FEB-1997  
ATTORNEY/AGENT INFORMATION:

REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7093

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441  
TELEX:  
INFORMATION FOR SEQ ID NO. 10.

SEQUENCE CHARACTERISTICS:  
LENGTH: 273 amino acids

TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
09-016-366A-19

Very Match	37.4%	Score 587.5	RD 3	1-1-11	666
------------	-------	-------------	------	--------	-----

[illegible][illegible]

9 LLLLCFSGORAKATACGRPRMLNRWVGODTQEGEMFWQVSIO---SHFCGSLI 65  
DB 3 LLLALPVLASRAAYAAPAPGALQVIGGQEARSKRPWQVSLRVHGFPMHFCGSLI 62  
QY 66 ABOVWTLAHCFF-RNTSETSLYOVLLGAROLVOPGHAMAYARVQVESNPLXOGTASSAD 124  
DB 63 HPQWVTLAHCFFGPDVDAALRVQLREOHLXY---ODOLLPVSRILVHPQFYTAIGAD 119  
QY 125 VALVELEAPVPTNYILPVCLPDPVSIFETGNCWYTGSGSPSEBLLPEPRILQKLAAP 184  
DB 120 IALLELEEVVNSSHVHTVTLPPASSETFPPGMCWYTGWDVNDERLPPFPPLKQVKVP 179  
QY 185 IIDTPKCNLYSKDTEFGYQPKTIKNDMLCAGFEESKKACKGDSGGLVCLVGSWTLQA 244  
DB 180 IMENHICDAKHILGAYTGDDVRIYRDMCLCAG--NTRRDSGCGDSGGLVCKVNGTWLQA 237  
245 GVISMGECAQONRPGVYIRVTAHNNHIIIRLPK 278  
238 GVSMGECQAOPNRPGIYTRVYTLDMIHVYK 271

RESULT 8  
US-08-978-404B-14  
Sequence 14, Application US/08978404B  
Patent No. 5968782  
GENERAL INFORMATION:  
APPLICANT: Stevens, Richard L.  
TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
NUMBER OF SEQUENCES: 74  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02210-2211  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/978,404B  
FILING DATE: 25-NOV-97  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/032,354  
FILING DATE: 04-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Plumer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7090  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
TELEX:  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 273 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: NO. 5968782e  
US-08-978-404B-14

Query Match 37.4%; Score 587.5; DB 2; Length 273;  
Best Local Similarity 43.8%; Pred. No. 4.9e-57;  
Matches 120; Conservative 46; Mismatches 99; Indels 9; Gaps 4;  
QY 9 LLLLCFSGORAKATACGRPRMLNRWVGODTQEGEMFWQVSIO---SHFCGSLI 65

DB 3 LLLALPVLASRAAYAAPAPGALQVIGGQEARSKRPWQVSLRVHGFPMHFCGSLI 62  
QY 66 ABOVWTLAHCFF-RNTSETSLYOVLLGAROLVOPGHAMAYARVQVESNPLXOGTASSAD 124  
DB 63 HPQWVTLAHCFFGPDVDAALRVQLREOHLXY---ODOLLPVSRILVHPQFYTAIGAD 119  
QY 125 VALVELEAPVPTNYILPVCLPDPVSIFETGNCWYTGSGSPSEBLLPEPRILQKLAAP 184  
DB 120 IALLELEEVVNSSHVHTVTLPPASSETFPPGMCWYTGWDVNDERLPPFPPLKQVKVP 179  
QY 185 IIDTPKCNLYSKDTEFGYQPKTIKNDMLCAGFEESKKACKGDSGGLVCLVGSWTLQA 244  
DB 180 IMENHICDAKHILGAYTGDDVRIYRDMCLCAG--NTRRDSGCGDSGGLVCKVNGTWLQA 237  
245 GVISMGECAQONRPGVYIRVTAHNNHIIIRLPK 278  
238 GVSMGECQAOPNRPGIYTRVYTLDMIHVYK 271

RESULT 9  
US-09-016-366A-23  
Sequence 23, Application US/09016366A  
Patent No. 595431  
GENERAL INFORMATION:  
APPLICANT: Stevens, Richard L.  
TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02210-2211  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016,366A  
FILING DATE: January 30, 1998  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/037,090  
FILING DATE: 05-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Plumer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7093  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
TELEX:  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 267 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-016-366A-23

Query Match 36.8%; Score 578.5; DB 2; Length 267;  
Best Local Similarity 43.6%; Pred. No. 4.7e-56;  
Matches 116; Conservative 45; Mismatches 96; Indels 9; Gaps 4;  
QY 17 SORAKATACGRPRMLNRWVGODTQEGEMFWQVSIO---RNGSHFCGSLIABOVWTLA 73  
DB 5 ASRAIYAPAPGALQVIGGQEARSKRPWQVSLRVHGFPMHFCGSLIHPQWTLA 64







Accession	Gene	Protein	Species	Length	Weight	PI	PI2	PI3	PI4	PI5	PI6	PI7	PI8	PI9	PI10	PI11	PI12	PI13	PI14	PI15	PI16	PI17	PI18	PI19	PI20	PI21	PI22	PI23	PI24	PI25	PI26	PI27	PI28	PI29	PI30	PI31	PI32	PI33	PI34	PI35	PI36	PI37	PI38	PI39	PI40	PI41	PI42	PI43	PI44	PI45	PI46	PI47	PI48	PI49	PI50	PI51	PI52	PI53	PI54	PI55	PI56	PI57	PI58	PI59	PI60	PI61	PI62	PI63	PI64	PI65	PI66	PI67	PI68	PI69	PI70	PI71	PI72	PI73	PI74	PI75	PI76	PI77	PI78	PI79	PI80	PI81	PI82	PI83	PI84	PI85	PI86	PI87	PI88	PI89	PI90	PI91	PI92	PI93	PI94	PI95	PI96	PI97	PI98	PI99	PI100	PI101	PI102	PI103	PI104	PI105	PI106	PI107	PI108	PI109	PI110	PI111	PI112	PI113	PI114	PI115	PI116	PI117	PI118	PI119	PI120	PI121	PI122	PI123	PI124	PI125	PI126	PI127	PI128	PI129	PI130	PI131	PI132	PI133	PI134	PI135	PI136	PI137	PI138	PI139	PI140	PI141	PI142	PI143	PI144	PI145	PI146	PI147	PI148	PI149	PI150	PI151	PI152	PI153	PI154	PI155	PI156	PI157	PI158	PI159	PI160	PI161	PI162	PI163	PI164	PI165	PI166	PI167	PI168	PI169	PI170	PI171	PI172	PI173	PI174	PI175	PI176	PI177	PI178	PI179	PI180	PI181	PI182	PI183	PI184	PI185	PI186	PI187	PI188	PI189	PI190	PI191	PI192	PI193	PI194	PI195	PI196	PI197	PI198	PI199	PI200	PI201	PI202	PI203	PI204	PI205	PI206	PI207	PI208	PI209	PI210	PI211	PI212	PI213	PI214	PI215	PI216	PI217	PI218	PI219	PI220	PI221	PI222	PI223	PI224	PI225	PI226	PI227	PI228	PI229	PI230	PI231	PI232	PI233	PI234	PI235	PI236	PI237	PI238	PI239	PI240	PI241	PI242	PI243	PI244	PI245	PI246	PI247	PI248	PI249	PI250	PI251	PI252	PI253	PI254	PI255	PI256	PI257	PI258	PI259	PI260	PI261	PI262	PI263	PI264	PI265	PI266	PI267	PI268	PI269	PI270	PI271	PI272	PI273	PI274	PI275	PI276	PI277	PI278	PI279	PI280	PI281	PI282	PI283	PI284	PI285	PI286	PI287	PI288	PI289	PI290	PI291	PI292	PI293	PI294	PI295	PI296	PI297	PI298	PI299	PI300	PI301	PI302	PI303	PI304	PI305	PI306	PI307	PI308	PI309	PI310	PI311	PI312	PI313	PI314	PI315	PI316	PI317	PI318	PI319	PI320	PI321	PI322	PI323	PI324	PI325	PI326	PI327	PI328	PI329	PI330	PI331	PI332	PI333	PI334	PI335	PI336	PI337	PI338	PI339	PI340	PI341	PI342	PI343	PI344	PI345	PI346	PI347	PI348	PI349	PI350	PI351	PI352	PI353	PI354	PI355	PI356	PI357	PI358	PI359	PI360	PI361	PI362	PI363	PI364	PI365	PI366	PI367	PI368	PI369	PI370	PI371	PI372	PI373	PI374	PI375	PI376	PI377	PI378	PI379	PI380	PI381	PI382	PI383	PI384	PI385	PI386	PI387	PI388	PI389	PI390	PI391	PI392	PI393	PI394	PI395	PI396	PI397	PI398	PI399	PI400	PI401	PI402	PI403	PI404	PI405	PI406	PI407	PI408	PI409	PI410	PI411	PI412	PI413	PI414	PI415
-----------	------	---------	---------	--------	--------	----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

XX New peptides useful for diagnosis, prevention and treatment of cancer  
PT and immune disorders

PS Claim 1; Page 145-146; 193pp; English.

CC AAV73325-Y73389 are human transcriptional regulator molecule (HTRM)  
CC protein sequences. The HTRM protein and nucleotide sequences are useful  
CC for preventing or treating disorders associated with decreased expression  
CC or activity of HTRM which include cell proliferative disorders such as  
CC arteriosclerosis and cirrhosis; cancers including adenocarcinoma and  
CC leukemia; immune disorders such as AIDS, Addison's disease, diabetes  
CC mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus  
CC erythematosus, and myasthenia gravis; infections and trauma. Antagonists  
CC of the HTRM polypeptides are useful for treating or preventing disorders  
CC associated with increased expression or activity of HTRMs. HTRM  
CC polypeptides, their immunogenic fragments or oligopeptides are useful  
CC for screening libraries of compounds in drug screening techniques.  
CC Polynucleotides encoding HTRM are useful for blocking the transcription  
CC of mRNA and regulating gene function by modulating the activity of  
CC HTRM. Vectors expressing HTRM or agonists can also be used to prevent or  
CC treat disorder associated with decreased HTRM expression. Antibodies  
CC which specifically bind HTRM and polynucleotides encoding HTRM are useful  
CC for diagnosing disorders associated with the expression of HTRM. Nucleotide  
CC sequences encoding HTRM may be useful to generate hybridization probes  
CC useful in mapping the naturally occurring genomic sequence and to detect  
CC differences in gene sequences among normal, carrier and affected  
CC individuals. Using diagnostic assays, cancer can be detected prior to  
CC the appearance of clinical symptoms and thereby progression of cancer can  
CC be prevented by aggressive treatment or preventive measures.

XX Sequence 290 AA:

Query Match 100.0%; Score 1571; DB 21; Length 290;  
Best Local Similarity 100.0%; Pred. No. 1.3e-134;  
Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRAAVPLLLILFCGSRKAKATACGRPLMRMYGSDTQGEPMWQVSIQDNGSHFC 60  
DB 1 MRRAAVPLLLILFCGSRKAKATACGRPLMRMYGSDTQGEPMWQVSIQDNGSHFC 60  
QY 61 GGSLLAEQWVLTARACFRNTSETSLYQVLGAKQVOPGHMAYARVOESNPLXQGTGA 120  
DB 61 GGSLLAEQWVLTARACFRNTSETSLYQVLGAKQVOPGHMAYARVOESNPLXQGTGA 120  
QY 121 SSADVALVELAEAVPPTNYTLPPCLDPSPYTFETGNNCWVTGSGSEEDILPEPTIIOK 180  
DB 121 SSADVALVELAEAVPPTNYTLPPCLDPSPYTFETGNNCWVTGSGSEEDILPEPTIIOK 180  
QY 181 LAVPIIDTPKCNLLYSKDTFEGYQPTIRKNDMLCAGFEGSKKDACGSDGSPVLCVIGQS 240  
DB 181 LAVPIIDTPKCNLLYSKDTFEGYQPTIRKNDMLCAGFEGSKKDACGSDGSPVLCVIGQS 240  
QY 241 WLOAGVYISNGEGCARONRGCVYIRVTAHNNHRIIRPKIOFOPARIGGCK 290  
DB 241 WLOAGVYISNGEGCARONRGCVYIRVTAHNNHRIIRPKIOFOPARIGGCK 290

RESULT 2  
AAU12282

ID AAU12282 standard; Protein; 290 AA.

AC AAU12282;

DE 24-OCT-2001 (first entry)

XX Human PRO4327 polypeptide sequence.

XX Human secretory and transmembrane; PRO; mammalian; cancer; lung;  
KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;  
KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;

KW adipocyte; A-peptide; factor VIIA; gene therapy.

OS Homo sapiens.

PN WO200140466-A2.

PD 07-JUN-2001.

PF 01-DEC-2000; 2000WO-US32678.

PR 01-DEC-1999; 99WO-US28301.

PR 01-DEC-1999; 99WO-US28634.

PR 02-DEC-1999; 99WO-US28551.

PR 02-DEC-1999; 99WO-US28564.

PR 02-DEC-1999; 99WO-US28565.

PR 09-DEC-1999; 99US-0170262.

PR 16-DEC-1999; 99WO-US30095.

PR 20-DEC-1999; 99WO-US30911.

PR 30-DEC-1999; 99WO-US30999.

PR 06-JAN-2000; 2000WO-US31243.

PR 06-JAN-2000; 2000WO-US30227.

PR 11-FEB-2000; 2000WO-US30376.

PR 18-FEB-2000; 2000WO-US30365.

PR 18-FEB-2000; 2000WO-US04342.

PR 22-FEB-2000; 2000WO-US04414.

PR 24-FEB-2000; 2000WO-US04914.

PR 01-MAR-2000; 2000WO-US05004.

PR 20-MAR-2000; 2000WO-US05601.

PR 21-MAR-2000; 2000WO-US07377.

PR 30-MAR-2000; 2000WO-US08439.

PR 17-MAY-2000; 2000WO-US13705.

PR 22-MAY-2000; 2000WO-US14042.

PR 30-MAY-2000; 2000WO-US14941.

PR 02-JUN-2000; 2000WO-US15264.

PR 10-NOV-2000; 2000WO-US30873.

PA (GETH) GENENTECH INC.

PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W,

PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;

PI Smith V, Stewart JA, Tumas D, Watanabe CK, Wood WI, Zhang Z;

DR WPI; 2001-408281/43.

DR N-PSDB; AAS21354.

XX Isolated, secretory and transmembrane PRO polypeptide used to detect

PT other PRO polypeptides, link bioactive molecules to cells expressing

PT PRO polypeptides, and detect the presence of mammalian tumours e.g.

lung, breast, prostate, cervical

Claim 12; Fig 222; 813pp; English.

CC AAU12172-AAU12446 represent novel human secretory and transmembrane  
CC PRO polypeptides. The PRO polypeptides are useful to detect other  
CC PRO polypeptides, to link bioactive molecules to cells expressing  
CC PRO polypeptides, to modulate biological activities of cells expressing  
CC PRO polypeptides, and to detect the presence of mammalian lung, colon,  
CC breast, prostate, rectal, cervical or liver tumours by comparing PRO  
CC polypeptide expression in a cell sample to that in a control sample.  
CC Some of the 275 sequences are also useful to stimulate the release of  
CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the  
CC proliferation or differentiation of chondrocytes, the proliferation or  
CC gene expression in pericyte cells, the release of proteoglycans from  
CC cartilage, the proliferation of inner ear utricular supporting cells or  
CC of T-lymphocytes, the release of a cytokine from peripheral blood  
CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of  
CC the PRO polypeptides may modulate glucose or free fatty acid uptake by  
CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide  
CC to factor VIIA. The PRO polypeptides can be used in assays to identify  
CC molecules involved in binding interactions. The polynucleotides encoding  
CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,



transgenic or knock out animals and can be used in gene therapy.

Sequence 290 AA:

Query Match 100.0%; Score 1571; DB 22; Length 290;  
Best Local Similarity 100.0%; Pred. No. 1.3e-134;  
Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRRPAVPLLLLCFSGSORAKATACGRPMINRMVGGDTQEGEMPWVSIQRNGSHFC 60  
1 mirpaavplllllcfsgsqrakatacgrpmInrmvsgdqtqegewpvsIqRngshfc 60  
61 GGSLLAEQWVLTAAHCFRNTSESLYQVLLGARQVOPPHMYARVQVESNPLYQGTGA 120  
61 ggsllaeqvwltahcfRntseSLyqVllgarqVlpphmyarVqvesnplyqgta 120  
121 SSADVALVELAEVPPFTNYLIPVCLDPPSVIFETGNMCWVTGWSPEEDLPEPRILQK 180  
121 ssadvalvelaevppftnylIPVclDppsvIFetgNmcwVtGwspeedllpeprIlqk 180  
181 LAVPIIDTPKCNLLYSKDTFEGYQPTIKNDMLCAGFEBSKDACGDSGGLVCLVYGQS 240  
181 lavpiidtpkcNllYSkdtfEGyqptIKndmLCagfeegkDacgdsGglvclVygqs 240  
241 WLQAGVTSNMGEGCARQNRPGVYTRVTAHNMWIRHRIIPKLOFOPARLGGOK 290  
241 wlqagvtsnmgEGcarQnrpgvYtrvTAhnmwIRhriIPkloFOpARlGGok 290

## RESULT 3

AAE03821  
ID AAE03821 standard; Protein; 290 AA.

AAE03821:

08-AUG-2001 (first entry)

Human gene 4 encoded secreted protein HMWH10, SRQ ID NO: 67.

Human; secreted protein; proliferative disorder; cancer; tumour; asthma;  
fetal abnormality; developmental abnormality; haematopoietic disorder;  
immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;  
psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;  
inflammation; neurological disorder; Alzheimer's disease; food additive;  
pregnancy-related disorder; kidney disorder; gastrointestinal disorder; allergy;  
cell culture; chemotaxis; vulnery; binding partner identification;  
gene therapy.

Homo sapiens.

Key Location/Qualifiers

Peptide 1..22

Protein /label= Signal\_peptide  
23..290  
/note= "Mature secreted protein"

WO200136440-A1.

25-MAY-2001.

15-NOV-2000; 2000WO-US31282.

19-NOV-1999; 99US-0166414.

21-JUL-2000; 2000US-0219665.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Komatsoulis GA, Birse CE, Moore PA;

WPI; 2001-343795/36.

DR N-PSDB: AAD08286.

Isolated nucleic acid molecule encoding a human secreted protein is  
used in preventing, treating or ameliorating a medical condition -

Claim 11; Page 482-483; 553pp; English.

AAD08283-AAD08355 represent cDNAs corresponding to 23 human secreted  
protein genes, and AAE03818-AAE03870 represent the proteins they encode.  
AAE03871-AAE03896 represent human secreted protein fragments or variants.  
The secreted proteins and their genes are useful for preventing,  
treating or ameliorating medical conditions, e.g., by protein or gene  
therapy. Pathological conditions can be diagnosed by determining the  
amount of the new protein in a sample or by determining the presence of  
mutations in the new genes. Specific uses are described for each of the  
23 genes, based on the tissues in which they are most highly expressed,  
and include developing products for the diagnosis or treatment of  
proliferative disorders, cancer, tumours, foetal and developmental  
abnormalities, haematopoietic disorders, diseases of the immune system,  
AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
allergies, neurological disorders (e.g., Alzheimer's disease,  
Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
cardiovascular disorders, angiogenic disorders, kidney disorders,  
gastrointestinal disorders, pregnancy-related disorders, endocrine  
disorders, and infections. The proteins can also be used to aid wound  
healing and epithelial cell proliferation, to prevent skin aging due to  
sunburn, to maintain organs before transplantation, for supporting cell  
culture of primary tissues, to regenerate tissues, to identify their  
cognate ligands or binding partners, and in chemotaxis, and can be used  
as a food additive or preservative to modify storage properties.  
Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
secreted protein of the invention.

Sequence 290 AA:

Query Match 100.0%; Score 1571; DB 22; Length 290;  
Best Local Similarity 100.0%; Pred. No. 1.3e-134;  
Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRRPAVPLLLLCFSGSORAKATACGRPMINRMVGGDTQEGEMPWVSIQRNGSHFC 60  
1 mirpaavplllllcfsgsqrakatacgrpmInrmvsgdqtqegewpvsIqRngshfc 60  
61 GGSLLAEQWVLTAAHCFRNTSESLYQVLLGARQVOPPHMYARVQVESNPLYQGTGA 120  
61 ggsllaeqvwltahcfRntseSLyqVllgarqVlpphmyarVqvesnplyqgta 120  
121 SSADVALVELAEVPPFTNYLIPVCLDPPSVIFETGNMCWVTGWSPEEDLPEPRILQK 180  
121 ssadvalvelaevppftnylIPVclDppsvIFetgNmcwVtGwspeedllpeprIlqk 180  
181 LAVPIIDTPKCNLLYSKDTFEGYQPTIKNDMLCAGFEBSKDACGDSGGLVCLVYGQS 240  
181 lavpiidtpkcNllYSkdtfEGyqptIKndmLCagfeegkDacgdsGglvclVygqs 240  
241 WLQAGVTSNMGEGCARQNRPGVYTRVTAHNMWIRHRIIPKLOFOPARLGGOK 290  
241 wlqagvtsnmgEGcarQnrpgvYtrvTAhnmwIRhriIPkloFOpARlGGok 290

## RESULT 4

AAE03945  
ID AAE03945 standard; Protein; 290 AA.

AAE03945:

29-MAY-2001 (first entry)

DE Human protease T.  
 XX  
 KW Human: protease T; serine protease; dermatological; desquamation;  
 KW skin care; laundry; detergent; shampoo; skin flaking.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200116293-A2.  
 XX  
 PD 08-MAR-2001.  
 XX  
 PF 30-AUG-2000; 2000WO-US23823.  
 XX  
 PR 31-AUG-1999; 99US-0386653.  
 XX  
 PA (ORTH ) ORTHO-MCNEIL PHARM INC.  
 XX  
 PI Darrow AL, Qi J, Andrade-Gordon P;  
 XX  
 PP WPI: 2001-265889/27.  
 XX  
 PS -PSDB; AAF76994.  
 XX  
 PT New serine protease termed protease T, useful for treating and  
 XX preventing skin flaking or imbalance of desquamation -  
 PS Claim 11; Fig 1; 83pp; English.  
 XX  
 CC The present sequence is human protease T protein. The protein is useful  
 CC for treating a condition mediated by protease T. It is useful for  
 CC treating an imbalance of desquamation, by topical application of  
 CC a pharmaceutical composition containing the protease. The composition is  
 CC useful as a topical skin care composition. It is useful as a laundry  
 CC detergent, shampoo, hard surface cleaning composition, and dish care  
 CC cleaning composition. Protease T protein is useful for treating and  
 CC preventing skin flaking. It is less immunogenic to sensitive individuals  
 CC and it provides efficient proteolytic activity in a non-natural  
 CC environment.  
 CC  
 CC Sequence 290 AA:  
 XX

Query Match 100.0%; Score 1571; DB 22; Length 290;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-134;  
 Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRPAVPLLLLCFSGORAKATACGRPMLNRMVGGQDTQEGEMPMOVSIOGNSHFC 60  
 DB 1 mrrpaavplllllcfsgqrakatacgrpmlnrmvvgdqtqegewpwwysiqngshfc 60

61 GGSIIAEQWVLTAAHCFRNTSETSLYQVLLGAROLVOPGPHAMARARQVESNPLVGGTA 120  
 61 ggsliieqvwltahcfrntsetsllyqvllygarqlvpgphamaryarvqesnplygta 120

QY 121 SSADVALVELLAPVPTNTYLLPVCLDPVYIPEFGMNCWTCGSPSEEDLPEPRLQK 180  
 DB 121 ssadvalvelleapvptntyllpvcldpvsyipefgmncwtcgspseedllpeprllqk 180

QY 181 LAPIIDTPKCNLLXKMDTEFGYQPTIKNDMLCAGFEESKKNACGSGGPIVCLVGOS 240  
 DB 181 lapiidtpkcncnllxkmdtefgyqptikndmlcagfeekknacgsggpiivclvgos 240

QY 241 WLQAGVISMEGECARONRPGVYIRVTAHHNMWIRITPKLOFOPARLGSGK 290  
 DB 241 wlqagviswegecaronrpgvyirvtahhnmwiritpklofoparlgsgk 290

RESULT 5  
 AAM39388  
 ID AAM39388 standard; Protein; 290 AA.  
 AC AAM39388;  
 XX  
 XX 22-OCT-2001 (first entry)  
 DT

XX  
 DE Human polypeptide SEQ ID NO 2533.  
 XX  
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW Leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US34263.  
 XX  
 PR 21-JAN-2000; 2000US-0488725.  
 XX  
 PR 25-APR-2000; 2000US-0552317.  
 XX  
 PR 09-JUL-2000; 2000US-0598042.  
 XX  
 PR 19-JUL-2000; 2000US-0620312.  
 XX  
 PR 03-AUG-2000; 2000US-0653450.  
 XX  
 PR 14-SEP-2000; 2000US-0662191.  
 XX  
 PR 19-OCT-2000; 2000US-0693036.  
 XX  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 XX  
 PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 XX  
 PI Zhao QA, Zhou P, Goodrich R, Dzmanac RT;  
 XX  
 DR WPI: 2001-442253/47.  
 XX  
 DR N-PSDB; AA158544.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 XX such as central nervous system injuries -  
 XX  
 PS Example 4; SEQ ID NO 2533; 10078pp; English.  
 XX

CC The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localized neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC assays for receptor activity, cancer diagnosis and therapy, drug screening,  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 CC  
 CC Sequence 290 AA:  
 XX

QY 1 MRRPAVPLLLLCFSGORAKATACGRPMLNRMVGGQDTQEGEMPMOVSIOGNSHFC 60  
 DB 1 mrrpaavplllllcfsgqrakatacgrpmlnrmvvgdqtqegewpwwysiqngshfc 60

QY 61 GGSIIAEQWVLTAAHCFRNTSETSLYQVLLGAROLVOPGPHAMARARQVESNPLVGGTA 120  
 DB 61 ggsliieqvwltahcfrntsetsllyqvllygarqlvpgphamaryarvqesnplygta 120

Query Match 99.6%; Score 1565; DB 22; Length 290;  
 Best Local Similarity 99.7%; Pred. No. 4.4e-134;  
 Matches 289; Conservative 0; Mismatches 1; Indels 0; Gaps 0;





RESULT 9  
 AAM8453  
 ID AAM8453 standard; Protein; 304 AA.  
 XX  
 AC AAM8453;  
 XX  
 DT 10-MAY-1999 (first entry)  
 XX  
 DE Serine protease HE2NM40.  
 XX  
 KM HE2NM40; serine protease; human; cancer; inflammation; asthma;  
 KM wasting disease; atherosclerosis; stroke; diabetes; arthritis;  
 KM neurodegenerative disease; Alzheimer's disease; diagnosis;  
 KM therapy; vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 PF EP890646-A2.  
 XX  
 PR 13-JAN-1999.  
 XX  
 PF 01-JUN-1998; 98EP-0304312.  
 XX  
 PR 20-FEB-1998; 98GB-0003650.  
 PR 10-JUN-1997; 97GB-0012088.  
 PR 17-OCT-1997; 97EP-0308295.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX  
 PI Burgess NA, Clinkenberg HE, Southan CD;  
 XX  
 DR WPI: 1999-072882/07.  
 DR N-PSDB: AAX06942.  
 XX  
 PT New serine protease HE2NM40 polypeptides and polynucleotides -  
 PT useful as diagnostic reagents and for prevention and treatment of  
 PT cancer, Alzheimer's disease and asthma  
 XX  
 PS Claim 1: Page 16-17; 21pp; English.  
 XX  
 CC This polypeptide, termed HE2NM40, comprises a serine protease that  
 CC that shows homology to human trypsin. Its amino acid sequence was  
 CC deduced from an isolated cDNA clone (see AAX06942). HE2NM40  
 CC polypeptides and polynucleotides are useful for diagnosing diseases  
 CC related to over or underexpression of HE2NM40 protein by identifying  
 CC mutations in the HE2NM40 gene, and/or analysing the presence or  
 CC amount of expressed polypeptide (claimed). HE2NM40 polypeptides,  
 CC polynucleotides (gene therapy), antisense sequences, agonists and  
 CC antagonists, and soluble polypeptides that bind the HE2NM40 ligand  
 CC to prevent expression, can be used to modulate HE2NM40 expression  
 CC or activity. Diseases diagnosed, prevented or treated include  
 CC cancer, inflammation, asthma, wasting diseases, atherosclerosis,  
 CC stroke, diabetes, arthritis, and neurodegenerative diseases and  
 CC disorders including Alzheimer's disease. HE2NM40 antibodies are  
 CC useful for inducing an immune response to immunise and prevent  
 CC disease, and for isolating HE2NM40 clones or purifying the  
 CC polypeptide by affinity chromatography. HE2NM40 polypeptides can  
 CC be administered directly or as a vaccine to inoculate against  
 CC disease.  
 XX  
 SQ Sequence 304 AA;  
 Query Match 76.28; Score 1197; DB 20; Length 304;  
 Best Local Similarity 82.14; Pred. No. 1.3e-100;  
 Matches 234; Conservative 7; Mismatches 36; Indels 8; Gaps 3;  
 QY 1 MRRRAAVPLLLLLCGSGORAKATACGRPTLNMVGGQPTQSEEMWQYSIQNGSHFC 60  
 DB 1 MRRRAAVPLLLLLCGSGORAKATACGRPTLNMVGGQPTQSEEMWQYSIQNGSHFC 60  
 QY 61 GGSIIAQWVLTAAHCRNMTSETSLYCVLLGARQLVPGHAMARVQVESNPLYGTA 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 61 ggsIIaeqwtlaahcfntsetslldqyllgarglvpgpghamyaryqvesnplygta 120  
 QY 121 SSADVATVLEAPVPEFTNYILPVCLPDPSVIFETGMNCWYMGSPSEEDLLPEPRILQ 180  
 DB 121 ssayvalveleapvpeftnyilpvclpdpsvifetgmncwytgmspsedllpeprilqk 180  
 QY 181 LAVPIIDPKCNLYSKDTEFGYQPKTKNDMLCAGFECKKADKDSGGLVCLVGS 240  
 DB 181 lavpidtpkcnylskdtetfgyqpktkndmlcagl-srrarmpasgagmwcsmgiq 238  
 QY 241 WLQAGV-LSWGEG-----CARQNRPGVYIRPTAHNMNTHRIPTL 279  
 DB 239 fgprpgpggvgispyrvwmqltpepsvllrvngqgmwlhrlmpxl 283  
 RESULT 10  
 AAY06482  
 ID AAY06482 standard; Protein; 317 AA.  
 XX  
 AC AAY06482;  
 XX  
 DT 27-SEP-1999 (first entry)  
 XX  
 DE Human tumour-associated protein PRO343.  
 XX  
 KM PRO343; UNQ302; cancer; tumour; diagnosis; therapy; human.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..32  
 FT Protein /note="signal peptide"  
 FT Protein 33..317  
 FT Modified-site /note="mature protein"  
 FT /note="N-glycosylated"  
 XX  
 PN WO935170-A2.  
 PD 15-JUL-1999.  
 PF 05-JAN-1999; 99MO-US00106.  
 XX  
 PR 20-NOV-1998; 98US-0109304.  
 PR 05-JAN-1998; 98US-0070440.  
 PR 22-APR-1998; 98US-0083500.  
 PR 22-MAY-1998; 98US-0086414.  
 PR 10-JUN-1998; 98US-0088742.  
 PR 10-NOV-1998; 98US-0107783.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Botstein D, Gôddard A, Gurney AL, Hillan KJ, Lawrence DA;  
 PI Roy MA, Wood WI;  
 XX  
 DR WPI: 1999-430385/36.  
 DR N-PSDB: AAX87259.  
 XX  
 PT Antibody against proteins expressed in neoplastic cells, useful for  
 PT tumor diagnosis and treatment  
 XX  
 PS Example 1; Fig 12; 162pp; English.  
 XX  
 CC This sequence represents human PRO343 (UNQ302), a protein encoded  
 CC by the novel cDNA clone DNA43318 (see AAX87259). Amplification of  
 CC DNA43318 was observed in primary lung tumours and in primary colon  
 CC tumours, suggesting a significant role in tumour formation and  
 CC growth. Antagonists (e.g. antibodies) directed to PRO343 may have  
 CC use in cancer therapy. The invention identifies 14 genes (see  
 CC AAX87254-67) that are amplified in the genome of tumour cells. Such  
 CC amplification is expected to be associated with overexpression of  
 CC the gene product and to contribute to tumorigenesis. The encoded  
 CC proteins (see AAY06477-90) may be useful targets for the diagnosis



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 08:56:23 ; Search time 30.25 Seconds  
(without alignments)  
1658.464 Million cell updates/sec

Title: US-10-041-006-7

Perfect score: 1571  
Sequence: 1 MRRPAAVPLLLLCRCGSGRA.....MIHRIPLKGFQPARLGQK 290

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SPREMBL\_19:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.mhc:\*  
9: sp.organelle:\*  
10: sp.phage:\*  
11: sp.plant:\*  
12: sp.rodent:\*  
13: sp.virus:\*  
14: sp.vertebrate:\*  
15: sp.unclassified:\*  
16: sp.rvivirus:\*  
17: sp.bacteriophage:\*  
18: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	721.5	45.9	297	11	088781 rat
2	682.5	43.4	389	13	09PVX7 xenopus lae
3	677.5	43.1	317	13	09DGR3 xenopus lae
4	633.5	40.3	310	11	091XC4 mus musculus
5	623.5	39.7	310	11	09QY29 mus musculus
6	612.5	39.0	321	4	096R28 mus musculus
7	610.5	38.9	339	11	099144 mus musculus
8	583.5	37.1	273	11	0921N4 mus musculus
9	573.5	36.5	275	4	096R26 mus musculus
10	555	35.3	799	11	09DB10 mus musculus
11	548	34.9	643	6	097506 mus musculus
12	545.5	34.7	273	6	09XSM1 ovine
13	542.5	34.5	329	13	042272 xenopus lae
14	525	33.4	237	6	046507 papio hamad
15	513.5	32.7	237	6	029464 bos taurus
16	505	32.1	806	6	018783 macropus eu

17	503	32.0	454	6	046506 papio hamad
18	503	32.0	810	4	015146
19	500.5	31.9	812	11	09R0M3
20	498.5	31.7	322	11	0920S2
21	494.5	31.5	490	11	0920K3
22	486.5	31.0	581	4	09BYE2
23	486	30.9	282	11	09DA13
24	485.5	30.9	624	6	095ME7
25	484	30.8	267	5	09BK47
26	483	30.7	624	11	09DA73
27	481.5	30.6	812	11	091W5
28	478.5	30.5	537	4	09BYE1
29	477	30.4	624	11	091Y47
30	471	30.0	492	4	096T73
31	469	29.9	1524	13	091674
32	467.5	29.8	405	4	096E86
33	462.5	29.4	855	11	093J17
34	462	29.4	329	6	096L10
35	455	29.0	479	5	0967X8
36	454.5	28.9	267	11	099MS4
37	453	28.8	326	11	09D9M0
38	452.5	28.8	263	11	09CR35
39	450.5	28.7	263	11	09DC86
40	449.5	28.6	845	13	09DGR1
41	448	28.5	279	11	09QZ74
42	448	28.5	415	6	029015
43	446.5	28.4	260	13	09W703
44	446.5	28.4	263	11	09DX8X
45	445.5	28.4	767	13	09DGR2

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	297 AA
1	088781			
AC	088781	PRELIMINARY	PRT	297 AA
DT	01-NOV-1998 (Tremblrel, 08, Created)			
DT	01-NOV-1998 (Tremblrel, 08, Last sequence update)			
DT	01-DEC-2001 (Tremblrel, 19, Last annotation update)			
DE	SERINE PROTEASE PRECURSOR (FRAGMENT).			
GN	BSP2.			
OS	Rattus rattus (Black rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10117;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FISHER; TISSUE=BRAIN;			
RC	MDLINE=98389725; PubMed=9722524;			
RA	Davies B.J., Pickard B.S., Steel M., Morris R.G., Lathe R.;			
RT	"Serine Proteases in Rodent Hippocampus.";			
RU	J. Biol. Chem. 273:23004-23011(1998)			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			
CC	TRYPSIN FAMILY.			
DR	EMBL; AJ005642; CAA0644.1; -			
DR	HSPB; P00763; IDPO.			
DR	MEROPS; S01.252; -			
DR	InterPro; IPR001314; Chymotrypsin.			
DR	InterPro; IPR001254; Trypsin.			
DR	Pfam; PF00899; Trypsin_1.			
DR	PRINTS; PR00722; CHYMOTRYPSIN.			
DR	SMART; SM00020; TRYP_SPC; 1.			
DR	PROSITE; PS00240; TRYPSIN_DOM; 1.			
DR	PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.			
DR	PROSITE; PS00135; TRYPSIN_SER; 1.			
KW	Hydrolase; Protease; Serine protease; Signal.			
FT	NON_TER			
FT	SIGNAL			
FT	CHAIN			
FT	SEQUENCE			

Query Match 45.9%; Score 721.5; DB 11; Length 297;  
 Best Local Similarity 48.3%; Pred. No. 2e-63;  
 Matches 18; Conservative 42; Mismatches 93; Indels 13; Gaps 5;

QY 4 PAAVPLLL--CFGSORAKAATACGRPMLNRYMGDGTGCEPMQVSIORNSHRC 60  
 DB 6 PGLTLEILPSATVSAANIRGSPDGKPKQNLNRYVGGEDSADQMPVYSLIKNSHHC 65  
 QY 61 GGSUAEQWVLTANCF-RMTSETSLYQVLLGARQLVDPGPHAMVAVRQVESNPLY-- 116  
 DB 66 AGSLNLRWVYSAHCFSSNMDKPSYVLGAMKLGNGFPRSQKVGASVLPHPKYSRK 125  
 QY 117 OCTASSADVALVELAPVPTNYLLPVCLPDPVETFGMCGVSGSPSEEDLLPEER 176  
 DB 126 EGT--HADIALVLERPIQESERILPICLPDSSVHLPPNTNCWIAQMSIQQGVLPKPO 183  
 QY 177 ILQKLVPIIDPKCNILYSKTEFGYQPKTKNDMLCAGFEKCKADCKSGGFLVCL 236  
 DB 184 TLQKLVPIIDPKCKSLYWR---GAGQEAILEDMLCAGLEGKRDACLDSGGFLMCO 239  
 QY 237 VQGSWLAGVISMGCACARONRPGVYIRYTAHNNHRIIPKLOPQ 282  
 DB 240 VDDHMLLTGIIISWEGCAERNRPGVTSLLAHPWQRTIVQGVQLR 285

RESULT 2  
 ID Q9PVX7 PRELIMINARY; PRT: 389 AA.  
 AC Q9PVX7:  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE EPIDERMIS SPECIFIC SERINE PROTEASE.  
 GN XEPSIN.  
 OS Xenopus laevis (African clawed frog).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 CC Xenopodinae; Xenopus.  
 RN NCBI\_TaxID=8355;  
 RP [1]  
 RA SEQUENCE FROM N.A.  
 RA Yamada K.;  
 RT "The expression control of xepsin by non-axial and planar  
 RT posteriorizing signals in Xenopus epidermis.";  
 RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 DB EMBL: AB018694; BAA84941.1; -  
 DB HSSP: P00763; IDPO.  
 DB MEROPS: S01.048; -  
 DB InterPro: IPR001314; Chymotrypsin.  
 DB InterPro: IPR001254; Trypsin.  
 DB Pfam: PF00089; Trypsin.1.  
 DB PRINTS: PR00722; CHYMOTRYPSIN.  
 DB SMART: SM00020; Tryp.sp.; 1.  
 DB PROSITE: PS00240; TRYPSIN\_DOM.1.  
 DB PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN.1.  
 DB PROSITE: PS00135; TRYPSIN\_SER.1.  
 KW Hydrolyase; Protease; Serine protease.  
 SQ SEQUENCE 389 AA; 42375 MW; B31FBA25D1F6E3 CRC64;

Query Match 43.4%; Score 682.5; DB 13; Length 389;  
 Best Local Similarity 48.3%; Pred. No. 2.1e-59;  
 Matches 125; Conservative 41; Mismatches 88; Indels 5; Gaps 3;

QY 25 ACGRPMLNRYMGDGTGCEPMQVSIORNSHRCGSLAEQWVLTANCFRMTSET 84  
 DB 16 ACGRPVLNRYMGDGTGCEPMQVSIORNSHRCGSLAEQWVLTANCFRMTSET 74  
 QY 85 LYQVLLGARQLVDPGPHAMVAVRQVESNPLYQGTASSADVALVELAPVPTNYLLPVC 144

DB 75 YTVYLLGAVQLSAPDNSTVSRGVKSTIKHPDFQYEGSSGDIALLEKPYFTPEYLLPIC 134  
 QY 145 LPDPVIFETGNCWVTGMSPEEDLLPEPRILQKLVPIIDPKCNILYSKTERGYQ 204  
 DB 135 LPSQVQFAAGIMCVTGMNIGQESTPLISPKTIQKAVAILIIDSVCSTMK--ESSIGYI 192  
 QY 205 P--KTIKNDMLCAGFEKCKADCKDSGGLVCLVQGSWLAGVISMGCACARONRPGV 262  
 DB 193 PDFSEFIQEDMVCAGYEGRIDACOGDSGGLVCNNVNWQLGLIVSWGCAEPRRPGV 252  
 QY 263 IRTVAHNNHRIIPKLOP 281  
 DB 253 TKVQYQDMKTNPLIVF 271

RESULT 3  
 ID Q9DGR3 PRELIMINARY; PRT: 317 AA.  
 AC Q9DGR3:  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DE EMBRYONIC SERINE PROTEASE-1.  
 GN XESP-1.  
 OS Xenopus laevis (African clawed frog).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 CC Xenopodinae; Xenopus.  
 RN NCBI\_TaxID=8355;  
 RP [1]  
 RA SEQUENCE FROM N.A.  
 RX MEDLINE=20363741; PubMed=10903452;  
 RA Yamada K.; Takabatake T.; Takeshima K.;  
 RT "Isolation and characterization of three novel serine protease genes  
 RT from Xenopus laevis.";  
 RL Gene 252:209-216(2000).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 DB EMBL: AB038496; BAB08216.1; -  
 DB HSSP: P00763; IDPO.  
 DB MEROPS: S01.048; -  
 DB InterPro: IPR001314; Chymotrypsin.  
 DB InterPro: IPR001254; Trypsin.  
 DB Pfam: PF00089; Trypsin.1.  
 DB PRINTS: PR00722; CHYMOTRYPSIN.  
 DB SMART: SM00020; Tryp.sp.; 1.  
 DB PROSITE: PS00240; TRYPSIN\_DOM.1.  
 DB PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN.1.  
 KW Hydrolyase; Protease; Serine protease.  
 SQ SEQUENCE 317 AA; 34413 MW; EEC78A9F46D138FE CRC64;

Query Match 43.1%; Score 677.5; DB 13; Length 317;  
 Best Local Similarity 46.8%; Pred. No. 5e-59;  
 Matches 123; Conservative 43; Mismatches 94; Indels 3; Gaps 2;

QY 22 AATACGRPMLNRYMGDGTGCEPMQVSIORNSHRCGSLAEQWVLTANCFRMTSET 81  
 DB 28 APPLCGSPVSSKIVGTDRQCAMPMQVSLKNSHRCGSLAEQWVLTANCFRMTSET 87  
 QY 82 ETSLYQVLLGARQLVDPGPHAMVAVRQVESNPLYQGTASSADVALVELAPVPTNYLL 141  
 DB 88 LPSGCGVRLGAYQLYKNPHEMTVKVDIYINSEFGPGTSGDIALKLSPIKFTPEYLL 147  
 QY 142 PVLQDPVIFETGNCWVTGMSPEEDLLPEPRILQKLVPIIDPKCNILYSKTERF 201  
 DB 148 PVLQDPVIFETGNCWVTGMSPEEDLLPEPRILQKLVPIIDPKCNILYSKTERF 207  
 QY 202 GYQPKTKNDMLCAGFEKCKADCKDSGGLVCLVQGSWLAGVISMGCACARONRPGV 261  
 DB 208 SEFEFIQEDMVCAGYEGRIDACOGDSGGLVCNNVNWQLGLIVSWGCAEPRRPGV 267  
 QY 262 YIRVTAHNNHRIIPKLOP 283



Query Match	39.0%;	Score 612.5;	DB 4;	Length 321;
Best Local Similarity	47.3%;	Pred. No. 1.5e-52;		
Matches 129;	Conservative 36;	Mismatches 91;	Indels 17;	Gaps 7

QY 9 LLLLCFSGORAKATACGRPMMLN---RMVGGODTQEGEMPMOVSTORNGSHFCGSLI 65  
 DB 9 LLLLAAGVGLRTLPQPCGRPOVSDAGRGVGHANAPAGAMPQASIRLRVAVCGSLL 68  
 QY 66 AEOVNLTAACFRMTSETSLYVLLGARQLVQRPAMARVROV--ESNPLOYGTASSA 123  
 DB 69 SPQWVLTAAHCFSGLNSSDYQVHLGELTTL-SPH--FSTVROILHSSPSGO-PGTS 124  
 QY 124 DVALVELAVPFTNYTLVPCLPDPSPVIFETGMNCWVTGSGSPSEEDLLPERILOKLAV 183  
 DB 125 DIALVELSVTVLSSRLTPCLPEASDDFCGIRCWVTGVTREGEPLRPYSLEKRV 184  
 QY 184 PIIDTPKCNLLYSKDEFGYQPKTKNDMLCAGFEKGKDKGSGGPLVCLVGSWQL 243  
 DB 185 SVYDTETCRDYP-----GPGGSILQPDMLCA---RGPQDAGDSDGSGPLVCVGNAMWQ 236  
 QY 244 AGVISMGECCARONRPGVYIRVTAHNMTHRIIT 276  
 DB 237 AGVISMGECCGRPNRPGVYIRVTAHNMTHRIIT 269  
 ID 099144 PRELIMINARY: PRT: 339 AA.  
 AC 099144: PRELIMINARY: PRT: 339 AA.  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE SIMILAR TO PROTEASE, SERINE, 8 (PROSTASIN).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV, AND SWISS; TISSUE=LUNG;  
 RA Verghese G.M., Caughey G.H.;  
 RT "Molecular cloning and characterization of mouse prostasin, a type I  
 RT membrane-associated serine protease of the gamma-tryptase/prostasin  
 RT gene family."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 DB EMBL: BC003851; AA03851.1; -  
 DB EMBL: AF378086; AAL06320.1; -  
 DB EMBL: AF378085; AAL06319.1; -  
 DB HSSP: P00734; 10VS.  
 DB MEROPS: S01.158; -  
 DB InterPro: IPR001314; Chymotrypsin.  
 DB InterPro: IPR001254; Trypsin.  
 DB Pfam: PF00089; trypsin.1.  
 DB PRINTS: PR00722; CHYMOTRYPSIN.  
 DB SMART: SM0020; TRYP-SPC; 1.  
 DB PROSITE: PS0240; TRYPSIN\_DOM; 1.  
 DB PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DB PROSITE: PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 339 AA; 36216 MW; BC2DE88BC057AF10 CRC64;

Query Match 38.9%; Score 610.5; DB 11; Length 339;  
 Best Local Similarity 42.3%; Pred. No. 2.5e-52;  
 Matches 119; Conservative 48; Mismatches 103; Indels 11; Gaps 4;

QY 6 AVPLLLILCF---GSQRAKATACG---RPMRLNRMVGGODTQEGEMPMOVSTORNGSHF 59  
 DB 14 AVTLLLLGLLQSGIRADGTREASCAVIOP---RITGGSAKPGQMPWVSYITDGNHV 69

QY 60 CGGSLAEQVNLTAACFRMTSETSLYVLLGARQLVQRPAMARVROVESNPLOYGT 119  
 DB 70 CGGSLVSNKMWVSAACHPEPREHREAREVKKGAHQDLSYNDTVVHTVAQIITHSYREE 129  
 QY 120 ASSAVVALVELEAVPFTNYTLVPCLPDPSPVIFETGMNCWVTGSGSPSEEDLLPERILQ 179  
 DB 130 GSGGIALIRLSSPTFSRTYRIPCLIPANASFPNGLCLVYTGWGHVAPSVLQTPRLQ 189  
 QY 180 KIAVDPIIDTPKCNLLYSKDEFGYQPKTKNDMLCAGFEKGKDKGSGGPLVCLVGSWQL 239  
 DB 190 QLEVLISRETSCSLYNINA--VPEEPHTIQDMLCAGYVKGKDKACGSDGSLSCPMEG 248  
 QY 240 SWLQGVISMGECCARONRPGVYIRVTAHNMTHRIIPKLO 280  
 DB 249 IWTLAGIVSWGDACAPNRPGVYITLTSTYASWIIHHVAELQ 289  
 ID 0921N4 PRELIMINARY: PRT: 273 AA.  
 AC 0921N4: PRELIMINARY: PRT: 273 AA.  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE MAST CELL PROTEASE 7.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC011328; AA011328.1; -  
 KW Protease.  
 SQ SEQUENCE 273 AA; 30332 MW; 51417476514035BE CRC64;

Query Match 37.1%; Score 583.5; DB 11; Length 273;  
 Best Local Similarity 45.1%; Pred. No. 8.9e-50;  
 Matches 123; Conservative 42; Mismatches 99; Indels 9; Gaps 5;

QY 10 LLLLCFSGORAKATACGRPMMLNRMVGGODTQEGEMPMOVSTORNGSHFCGSLI 66  
 DB 4 LLLTLPLLSLVLAAGPAMTREGIVGQEAHGNKMPWVSLRANDYTMHFCGSLIH 63  
 QY 67 EOWVLTAAHCF-RNTSETSLYVLLGARQLVQRPAMARVROVESNPLOYGTASSADY 125  
 DB 64 POWVLTAAHCVGRVADPNKVRVQLRKQYLY--HDHLMTVSQIITHPDFYIVQDADI 120  
 QY 126 ALVELAVPFTNYTLVPCLPDPSPVIFETGMNCWVTGSGSPSEEDLLPERILOKLAVP 185  
 DB 121 ALKLTMPVAVISDYVHVRLPRASETPPSGTLQWVGWINDGVNLPPPPKKEVQVPI 180  
 QY 186 IDPKCNLLYSKDEFGYQPKTKNDMLCAGFEKGKDKGSGGPLVCLVGSWQL 245  
 DB 181 IENHLDLKYHKGILITGDVNIYVDDMLCAG-NEG-HDSCGDSGSGPLVCXVEDTMDLAG 238  
 QY 246 VISMGECARONRPGVYIRVTAHNMTHRIIPK 278  
 DB 239 VISMGECCAPNRPGVYITVYLYLDWIMHRYVR 271

RESULT 9  
 ID 096RZ6 PRELIMINARY: PRT: 275 AA.  
 AC 096RZ6: PRELIMINARY: PRT: 275 AA.  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE TRYPTASE 1.  
 GN TRYPTASE.  
 OS Homo sapiens (Human).

RA  
sasaki H., Sato K., Schoenbach C., seya T., Shlbata Y., Storch K.-F.,

QY	26	CGPRRLINAMWGQDZTGGEMPMOVSITORNGSHCCGSLIAEQMVTLANHCFRNTSFRS	84
Db	556	CGIAGLSSTAVGGYSSGEGENPMQASLIQIRNHCIGALLIDRNMVTIANHCCQEDSMSP	6151
QY	85	-LYVLLGARQLVORCPAHMYARVROYESNPELYOGTASADVALVELEAPVEFTNYILPV	143
Db	616	KLMTVELTGLKRONSRBMPGEVSFKYSRLFLRPHNEEDSHDVALQLDLHPVYVATVPV	6757
QY	144	CLRDPEVIREPTGNOMVWGMSPSEBDLLEPRITLQKLAIPIDDPKCLYSKNTERGY	203
Db	676	CLARSHFFERQCHMWTGWSQREBG--PVSNTLQAVDVQVLPDQLCSAYR-----Y	722
QY	204	QPTIRKNDMLCAGFEFGKDKADQSGGRLVLCVGS--WLAGVISMEGCARONRPGV	263
Db	728	Q---VSPRMLCAGYRKKDKADQSGSGRLVCRPSGWRFLAGVSMGLGCGRPNFGVY	784
QY	263	IRYTAHHNNIIRRI	276
Db	785	IRVTRVIMNIQVLL	798

Search completed: August 13, 2002, 09:00:31  
Job time: 248 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: August 13, 2002, 08:57:03 ; Search time 13.47 Seconds  
(without alignments)  
833.606 Million cell updates/sec

Title: US-10-041-006-7

Perfect score: 1571  
Sequence: 1 MRRAVAVPLLLLCFSGSRA.....WIRRIPLKQFQPARLGSGQ 290

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1571	100.0	290	1 MPN_HUMAN	O9pqr3 homo sapien
2	739	47.0	317	2 BSS4_HUMAN	O9gzn4 homo sapien
3	696	44.3	306	1 BSS4_MOUSE	O9er10 mus musculu
4	618.5	39.4	343	1 PSS8_HUMAN	O16651 homo sapien
5	614.5	39.1	342	1 PSS8_RAT	O9es87 rattus norv
6	613.5	39.1	321	1 TRYG_HUMAN	O9nrr2 homo sapien
7	610.5	38.9	311	1 TRYG_MOUSE	O9qu17 mus musculu
8	606.5	38.6	342	1 PSS8_MOUSE	O9esd1 mus musculu
9	594	37.8	270	1 TRYT_MERUN	P50342 meriones un
10	590.5	37.6	273	1 MCT7_RAT	P20231 rattus norv
11	588.5	37.5	275	1 TRB1_HUMAN	O15661 homo sapien
12	587.5	37.4	275	1 TRYT_HUMAN	O9nrd1 sus scrofa
13	584.5	37.2	273	1 MCT7_MOUSE	O9ed44 mus musculu
14	578.5	36.8	273	1 TRYT_MOUSE	O9yem0 homo sapien
15	574	36.5	314	1 TRYT_HUMAN	P15157 homo sapien
16	571.5	36.4	275	1 TRYA_HUMAN	P50343 rattus norv
17	565.5	36.0	274	1 MCT6_RAT	P21845 mus musculu
18	564	35.9	276	1 MCT6_MOUSE	P15944 canis faml
19	560.5	35.7	275	1 TRYT_CANFA	O9jnj7 mus musculu
20	559.5	35.6	324	1 TEST_MOUSE	O9xsm2 ovis aries
21	556.5	35.4	273	1 TRYT_SHEEP	O9xsm2 ovis aries
22	533	33.9	638	1 KAL_HUMAN	P12545 macaca mula
23	526	33.5	810	1 PLMN_MACMU	O35453 mus musculu
24	523.5	33.3	416	1 HEPS_MOUSE	P26262 mus musculu
25	518	33.0	638	1 KAL_MOUSE	P05981 homo sapien
26	517.5	32.9	417	1 HEPS_HUMAN	P14272 rattus norv
27	517	32.9	638	1 KAL_RAT	P80010 equus cabal
28	509	32.4	338	1 PLMN_HORSE	O05511 rattus norv
29	505.5	32.2	416	1 HEPS_RAT	O9er10 mus musculu
30	504.5	32.1	455	1 TMS5_MOUSE	P00747 homo sapien
31	502	32.0	810	1 PLMN_HUMAN	P06868 bos taurus
32	496.5	31.6	333	1 PLMN_BOVIN	P80009 canis faml
33	496	31.6	333	1 PLMN_CANFA	

34	493	31.4	269	1 TRYM_CANFA	P19236 canis faml
35	491.5	31.3	1034	1 ENTK_PIG	P98074 sus scrofa
36	488.5	31.1	450	1 TMS2_MOUSE	O91j98 mus musculu
37	487.5	31.0	625	1 FALL_HUMAN	P03951 homo sapien
38	485.5	30.9	457	1 TMS2_HUMAN	O9n383 homo sapien
39	481	30.6	422	1 DES1_HUMAN	O9u152 homo sapien
40	479.5	30.5	812	1 PLMN_MOUSE	P20918 mus musculu
41	479	30.5	790	1 PLMN_PIG	P06867 sus scrofa
42	478	30.4	343	1 PLMN_SHEEP	P14417 macaca mula
43	477.5	30.4	1420	1 APON_MACMU	P81286 ovis aries
44	476.5	30.3	855	1 ST14_HUMAN	O9y5y6 homo sapien
45	471	30.0	492	1 TMS2_HUMAN	O13593 homo sapien

## ALIGNMENTS

RESULT	ID	MPN_HUMAN	STANDARD	PRT	290 AA.
AC	O9BOR3	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Marapsin precursor (EC 3.4.21.-)				
GN	MPN				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Fortunato M., Dando P.M., Rawlings N.D., Barrett A.J.;				
RT	"Cloning, sequencing and expression of marapsin, a human serine				
RL	protease."				
CC	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.				
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE				
CC	TRYPsin FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement. (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
CC	EMBL: AJ306593; CAC35467.1; -				
CC	MEPROS; S01.074; -				
CC	InterPro: IPR001314; Chymotrypsin.				
CC	InterPro: IPR001234; Trypsin.				
CC	Pfam: PF00089; Trypsin; 1.				
CC	PRINTS: PR00722; CHYMOTRYPSIN.				
CC	SMART: SM00020; TRYP-SPC; 1.				
CC	PROSITE: PS0240; TRYPsin; DOM; 1.				
CC	PROSITE: PS00134; TRYPsin; HIS; 1.				
CC	PROSITE: PS00135; TRYPsin; SER; 1.				
CC	Hydrolyase; Serine protease; Lysogen; Signal; Glycoprotein.				
CC	SIGNAL	1	22		
CC	POTENTIAL.				
CC	PROPEP	23	34		
CC	CHAIN	35	290		
CC	DOMAIN	35	277		
CC	ACT_SITE	75	75		
CC	ACT_SITE	124	124		
CC	ACT_SITE	229	229		
CC	DISULFID	60	76		
CC	DISULFID	158	235		
CC	DISULFID	191	214		
CC	DISULFID	225	253		
CC	CAROHWD	55	55		
CC	CAROHWD	79	79		
CC	SEQUENCE	290 AA;	31940 MW;	67BDC93BC70BFF7B CRC64;	

Query Match 100.0%; Score 1571; DB 1; Length 290;  
 Best Local Similarity 100.0%; Pred. No. 1,3e-135;  
 Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRPAAPVLLLLLCFGSGORAKAATACGRPMILNRMYGQDTPGEMPMOVSITORNSHFC 60  
 DB 1 MRPAAPVLLLLLCFGSGORAKAATACGRPMILNRMYGQDTPGEMPMOVSITORNSHFC 60  
 QY 61 GGSIIAEQWVLTAAHCRNRTSETSLVQVLLGAQVOPGPHAMTARVQESNPLYQSTA 120  
 DB 61 GGSIIAEQWVLTAAHCRNRTSETSLVQVLLGAQVOPGPHAMTARVQESNPLYQSTA 120  
 QY 121 SSADVALVELAPVFNITLPCVLPDPSPVIFETGMNCWVTGWSGSPSEEDLPEPRILQK 180  
 DB 121 SSADVALVELAPVFNITLPCVLPDPSPVIFETGMNCWVTGWSGSPSEEDLPEPRILQK 180  
 QY 181 LAVIITDPKCNLTYSKDFEQPTKINDMLCAGFEKGKDKGSGGLVCLVQSGS 240  
 DB 181 LAVIITDPKCNLTYSKDFEQPTKINDMLCAGFEKGKDKGSGGLVCLVQSGS 240  
 QY 241 WLAGVISMEGGCARONRPGVYIRVTAHNMWTHRIIPKLOFOPARLGQK 290  
 DB 241 WLAGVISMEGGCARONRPGVYIRVTAHNMWTHRIIPKLOFOPARLGQK 290

## RESULT 2

BSS4\_HUMAN STANDARD; PRT; 317 AA.

AC 09GZM4; 043342;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Brain-specific serine protease 4 precursor (EC 3.4.21.-) (BSSP-4) (SP001LA).  
 GN PRSS26 OR PRSS22 OR BSSP4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Mitsui S., Okui A., Komiami K., Yamaguchi N.;  
 RT "Cloning and characterization of a human brain-specific serine protease, hbssp-4.";  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RA Young G.W., Stevens R.L.;  
 RT Identification of a new member of the chromosome 16 family of serine proteases.";  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

SEQUENCE OF 47-317 FROM N.A.  
 R1 R1C6 D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,  
 RA Robinson D., Jones M., Buckingham J., Chastain L., Thompson S.,  
 RA Goodwin L., Bryant J., Tesmer J., Melnick L., Longmire J., White S.,  
 RA Deng S., Tatum O., Campbell C., Fawcett J., Deaven L.;  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.

-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL, AB010779; BAB20263.1; -

DR EMBL; AF321182; AAG35070.1; -;  
 DR EMBL; AC003965; AAB93671.1; -;  
 DR MEROPS; S01.252; -;  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Trypsin.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYP-Spc; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM Hydrolyase; Serine protease; Signal.  
 FT SIGNAL 1 32  
 FT CHAIN 33 317  
 FT ACT\_SITE 90 90  
 FT ACT\_SITE 141 141  
 FT ACT\_SITE 242 242  
 FT DISULFID 75 91  
 FT DISULFID 175 248  
 FT DISULFID 208 227  
 FT DISULFID 238 266  
 FT CARBOHYD 70 70  
 FT CONFLICT 47 47  
 SQ SEQUENCE 317 AA; 33731 MW; E2A123BC6E79935 CRC64;

Query Match 47.0%; Score 739; DB 1; Length 317;  
 Best Local Similarity 48.8%; Pred. No. 6,3e-60;  
 Matches 139; Conservative 49; Mismatches 87; Indels 10; Gaps 4;

QY 10 LLLLC----FGSRAKAAATACGRPMILNRMYGQDTPGEMPMOVSITORNSHFCGSLI 65  
 DB 21 LLLLCSTALNAAKLPVPPACGKPDQNLNVGSDSTSEPMVVSIOKNGTHHAGSLI 80  
 QY 66 AQQWVLTAAHCRN-NTSETSLVQVLLGAQVOPGPHAMTARVQESNPLYQ-GRASSA 123  
 DB 81 TSNWVITAAHCEKDLNLPVLLGAMQNGSRSQKVGAVNPPHVVYWKAGACA 140  
 QY 124 DVATVELAPVFNITLPCVLPDPSPVIFETGMNCWVTGWSGSPSEEDLPEPRILQKLV 183  
 DB 141 DVATVELAPVFNITLPCVLPDPSPVIFETGMNCWVTGWSGSPSEEDLPEPRILQKLV 183  
 QY 184 PIIDPKCNLTYSKDFEQPTKINDMLCAGFEKGKDKGSGGLVCLVQSGSLQ 243  
 DB 201 PIIDSEVCSHLYWAGAGG-----PIIDMLCAGTLEBERACLDGSGGLVCLVQSGSLQ 256  
 QY 244 AGVISMEGGCARONRPGVYIRVTAHNMWTHRIIPKLOFOPARLG 288  
 DB 257 AGVISMEGGCARONRPGVYISLSAHSRWEKIVQGVQLRGAQGG 301

## RESULT 3

BSS4\_MOUSE STANDARD; PRT; 306 AA.

AC 09ERIO;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Brain-specific serine protease 4 precursor (EC 3.4.21.-) (BSSP-4).  
 GN Mus musculus (Mouse).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Mitsui S., Okui A., Komiami K., Yamaguchi N.;  
 RT "Cloning and characterization of a novel serine protease, mbssp-4.";  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL: AB010778; BAB20262.1; -  
 CC InterPro: IPR001314; Chymotrypsin.  
 CC InterPro: IPR001254; Trypsin.  
 CC Pfam: PF00089; trypsin.1  
 CC PRINTS: PR00722; CHYMOTRYPSIN.  
 CC SMART: SM00020; TRYP\_SPE; 1.  
 CC PROSITE: PS50240; TRYPsin\_DOM; 1.  
 CC PROSITE: PS00134; TRYPsin\_HIS; 1.  
 CC PROSITE: PS00135; TRYPsin\_SER; 1.  
 CC Hydrolase: Serine protease; Signal.  
 CC SIGNAL 1 32 POTENTIAL  
 CC CHAIN 33 306 BRAIN-SPECIFIC SERINE PROTEASE 4.  
 CC ACT\_SITE 90 90 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC ACT\_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC ACT\_SITE 242 242 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC DISULFID 75 91 BY SIMILARITY.  
 CC DISULFID 175 248 BY SIMILARITY.  
 CC DISULFID 208 227 BY SIMILARITY.  
 CC DISULFID 238 266 BY SIMILARITY.  
 CC CARBOHYD 70 70 N-LINKED (GLCNAC... ) (POTENTIAL).  
 CC SEQUENCE 306 AA; 33262 MW; FBBF03C0C285E7E8 CRC64;

Query Match 44.38; Score 696; DB 1; Length 306;  
 Best Local Similarity 47.68; Pred. No. 4.9e-56;  
 Matches 136; Conservative 47; Mismatches 87; Indels 16; Gaps 6;

OY 6 AVPLLLLCFSSORAKATA-----CGRPMILNRMVGGQDQOEEMPMQVSIORNGSHFC 60  
 DB 16 SLILLVLTSTAPISATINIVSPCGKQDLNRTVGSEDMQMPVIVSLKNGSHHC 75  
 OY 61 GGLIAEQWVLTAAHCFR-NTSEISLYVLLGAROLVOPGPHAMARYQVESNPLY--- 116  
 DB 76 AGSLTNMNVVYAAHCFKSNMCKPFLFGLGAMKLGSPGPRQVGIWVLPHPRYSWK 135  
 OY 117 OCTASSADVAVLEAPVPTNYILPVCLPSPVFEFGNMQWVGWSPSEEDLPPPR 176  
 DB 136 EST--HADIALVRLHSHISQFSERILPCLPSSVVALPKTKDCLWAGWSIDGCVLPHPQ 193  
 OY 177 ILQKLVPIIDTPKCNLLYSKDFEYQPKTKINDMLCAGEEGKKDCKDGGGPIVCL 236  
 DB 194 TLQKLVPIIDSELCISLYWR---GAGQEAITEGMLCAGYLEGRDACLGDGSGPLMCQ 249  
 OY 237 VQOSMLQACVSIWEGCAGQRNRPVYIRVTAHNNHRIKPLQFQ 282  
 DB 250 VDDHMLTLGIISWEGCA-DDRPGVYTSILAHRSVQRIYQGVOLR 294

RESULT 4  
 PSS8\_HUMAN STANDARD; PRT; 343 AA.  
 AC Q16651; Q9UCA3;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Protaasin precursor (EC 3.4.21.-).  
 GN PRSS8.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Prostate;  
 RX MEDLINE=95286644; PubMed=7768952;

RA Yu J.X., Chao L., Chao J.;  
 RT "Molecular cloning, tissue-specific expression, and cellular  
 RT localization of human prostaasin mRNA."  
 RL J. Biol. Chem. 270:13483-13489(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Strausberg R.;  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 45-64.  
 RC TISSUE=semen;  
 RX MEDLINE=94308140; PubMed=8034638;  
 RA Yu J.X., Chao L., Chao J.;  
 RT "Protaasin is a novel human serine proteinase from seminal fluid.  
 RT Purification, tissue distribution, and localization in prostate  
 RT gland.";  
 RL J. Biol. Chem. 269:18843-18848(1994).  
 CC -1- FUNCTION: POSSESSES A TRYPSIN-LIKE CLEAVAGE SPECIFICITY.  
 CC -1- SUBUNIT: HETERODIMER OF TWO CHAINS, LIGHT AND HEAVY, HELD BY A  
 CC DISULFIDE BOND.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF  
 CC ITS C-TERMINUS.  
 CC -1- TISSUE SPECIFICITY: FOUND IN PROSTATE, LIVER, SALIVARY GLAND,  
 CC KIDNEY, LUNG, PANCREAS, COLON, BRONCHUS AND RENAL PROXIMAL TUBULAR  
 CC CELLS. IN THE PROSTATE GLAND IT MAY BE SYNTHESIZED IN EPITHELIAL  
 CC CELLS, SECRETED INTO THE DUCTS, AND EXCRETED INTO THE SEMINAL  
 CC FLUID.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L41351; AAC41759.1; -  
 DR EMBL: U33446; AAB19071.1; -  
 DR EMBL: BC001462; AAB01462.1; -  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.159; -.  
 DR MIM: 600823; -  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin.1  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRYP\_SPE; 1.  
 DR PROSITE: PS50240; TRYPsin\_DOM; 1.  
 DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE: PS00135; TRYPsin\_SER; 1.  
 DR Hydrolase: Serine protease; Zymogen; Signal; Glycoprotein;  
 KW Transmembrane.  
 FT SIGNAL 1 29 POTENTIAL.  
 FT PROPEP 30 32 ACTIVATION PEPTIDE.  
 FT CHAIN 33 44 PROSTASIN LIGHT CHAIN.  
 FT CHAIN 45 322 PROSTASIN HEAVY CHAIN.  
 FT PROPEP 323 343  
 FT TRANSSEM 320 340  
 FT DOMAIN 45 286  
 FT DISULFID 37 154  
 FT DISULFID 70 86  
 FT DISULFID 168 244  
 FT DISULFID 201 223  
 FT DISULFID 234 262  
 FT ACT\_SITE 85 85  
 FT ACT\_SITE 134 134  
 FT ACT\_SITE 238 238  
 FT CARBOHYD 159 159  
 FT SEQUENCE 343 AA; 36431 MW; 98D06447F5A8C1B2 CRC64;

Query Match 39.4%; Score 618.5; DB 1; Length 343;  
 Best Local Similarity 42.9%; Pred. No. 6,2e-49;  
 Matches 124; Conservative 47; Mismatches 99; Indels 19; Gaps 7;

QY 6 AVPLLL--LLCFSGQRAKATACG--RPRMLNMGVGGDTQEGEMVQVSTORNGSHFCG 61  
 DB 14 AVAILLYLLGRLSGTAGEGAEAPCAPOA--RITGSSAVAGQMPWQVSTIREGVHVG 71  
 QY 62 GSIIAEQWVLTAAHCFRNTSETSLYOVLLGAROLVPGHMYARAROVESNLYGTGTS 121  
 DB 72 GSIVSQWVLTAAHCFRNTSETSLYOVLLGAROLVPGHMYARAROVESNLYGTGTS 131  
 QY 122 SADVALVELAPVFTNYLLPVCLPDPVIFETGNCWYTGNG--SPSEEDLLPEPRILQ 179  
 DB 132 QGDIALLOLSRPTFSRYRIPCLPANAASFPGHLCYTGWHVAPSVSLTPKPK--LQ 189  
 QY 180 KLAVALPIIDPKKLLSKTERGYOPTIKNDMLCAGEEGKADCKGDSGGLVCLVQ 239  
 DB 190 QLEVPILISRETCLYNIDAK--PEEPHVEQEDWVAGYVEGGKADCKGDSGGLVCLVQ 248  
 QY 240 SWLQAGVISMGECCARONRPGVYIRVTAHNMTH-----RIIPKIQ 280  
 DB 249 LWLTGIVSMGDACGARNRPGVYITLSTASVMSIQKVTLEDPVPPQ 297

RESULT 5  
 PSS8\_RAT STANDARD; PRT; 342 AA.

ID PSS8\_RAT STANDARD; PRT; 342 AA.  
 AC Q9ES87: Q9ER01; 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Proctasin precursor (EC 3.4.21.-).  
 GN PRSS8.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI-TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RA Adachi M., Kitamura K., Miyoshi T., Tomita K.;  
 RL Submitted (Sep-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Wang C.;  
 RT "Molecular cloning and expression of rat proctasin";  
 RI Submitted (Nov-1999) to the EMBL/GenBank/DBJ databases.  
 RI - FUNCTION: POSSESSES A TRYPSIN-LIKE CLEAVAGE SPECIFICITY (BY SIMILARITY).  
 CC - SUBUNIT: HETERODIMER OF TWO CHAINS, LIGHT AND HEAVY, HELD BY A DISULFIDE BOND (BY SIMILARITY).  
 CC - SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF ITS C-TERMINUS (BY SIMILARITY).  
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to [license@isb.sib.ch](mailto:license@isb.sib.ch)).

CC EMBL: AB017638; BAB20281.1; -  
 DR EMBL: AF202076; AAG32641.1; -  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF000089; trypsin.1.  
 DR SMART: SM00020; TRYP-SPC; 1.  
 DR PROSITE: PS02340; TRYPSIN\_DOM; 1.  
 KW Hydrolase; Serine protease; Zymogen; Signal; Glycoprotein;

Transmembrane.  
 KM SIGNAL 1 29  
 FT PROPEP 30 32  
 FT CHAIN 33 44  
 FT CHAIN 45 45  
 FT CHAIN 45 322  
 FT PROPEP 323 342  
 FT TRANSMEM 320 340  
 FT DOMAIN 45 286  
 FT DISULFID 37 154  
 FT DISULFID 70 86  
 FT DISULFID 168 244  
 FT DISULFID 201 223  
 FT DISULFID 234 262  
 FT ACT\_SITE 85 85  
 FT ACT\_SITE 134 134  
 FT ACT\_SITE 238 238  
 FT CARBOHYD 159 159  
 FT CONFLICT 17 17  
 FT CONFLICT 292 292  
 SQ SEQUENCE 342 AA; 36843 MW; 5ED1AF05D213B98 CRC64;

Query Match 39.1%; Score 614.5; DB 1; Length 342;  
 Best Local Similarity 42.5%; Pred. No. 1.4e-48;  
 Matches 119; Conservative 46; Mismatches 102; Indels 13; Gaps 4;

QY 9 LLLLCFSGQRAKATACG--RPRMLNMGVGGDTQEGEMVQVSTORNGSHFC 60  
 DB 15 LFIILLGLQSRIGADGTEASCAVLP--RITGSGAKPGQWQVSTIRNVCVHVC 70  
 QY 61 GSIIAEQWVLTAAHCFRNTSETSLYOVLLGAROLVPGHMYARAROVESNLYGTGTA 120  
 DB 71 GSIVSQWVLTAAHCFRNTSETSLYOVLLGAROLVPGHMYARAROVESNLYGTGTA 130  
 QY 121 SADVALVELAPVFTNYLLPVCLPDPVIFETGNCWYTGNG--SPSEEDLLPEPRILQ 180  
 DB 131 QGDIALLOLSRPTFSRYRIPCLPANAASFPGHLCYTGWHVAPSVSLTPKPK--LQ 190  
 QY 181 LAVPIIDPKKLLSKTERGYOPTIKNDMLCAGEEGKADCKGDSGGLVCLVQ 240  
 DB 191 LEVPIIDPKKLLSKTERGYOPTIKNDMLCAGEEGKADCKGDSGGLVCLVQ 249  
 QY 241 WLAGVISMGECCARONRPGVYIRVTAHNMTHRIIPKIQ 280  
 DB 250 WYLAGIVSMGDACGARNRPGVYITLSTASVMSIQKVTLEDPVPPQ 289

RESULT 6  
 TRIG\_HUMAN STANDARD; PRT; 321 AA.

ID TRIG\_HUMAN STANDARD; PRT; 321 AA.  
 AC Q9NR82: Q9NR08; Q9C015; Q9UBB2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE tryptase gamma precursor (EC 3.4.21.-) (Transmembrane tryptase).  
 GN TP5G1 OR TMT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI-TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (VARIANTS GAMMA-1 AND GAMMA-2).  
 RX MEDLINE-20302813; PubMed-10843716;  
 RA Caughey G.H., Raymond W.W., Blount J.L., Hau L.W., Pallao M.,  
 RA Wolters P.J., Verghese G.M.;  
 RT "Characterization of human gamma-tryptase, novel members of the  
 RT chromosome 16p mast cell tryptase and proctasin gene families";  
 RL J. Immunol. 164:6566-6575(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-99452974; PubMed-10521469;  
 RA Wong G.W., Tang Y., Feyfant E., Sali A., Li L., Li Y., Huang C.,  
 RA Friend D.S., Krilis S.A., Stevens R.L.;





CC  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC

AA MEDLINE-505009/1, FUDMED-1033

RX MEDLINE=95366971; PubMed=7639711;

RA Murakumo Y., Ide H., Itoh H., Tomita M., Kobayashi T.,  
 RA Maruyama H., Horii Y., Nawa Y.,  
 RT "Cloning of the cDNA encoding mast cell tryptase of Mongolian gerbil,  
 RT Meriones unguiculatus, and its preferential expression in the  
 RT intestinal mucosa." J. 309:921-926(1995).  
 RL Blochem. J. 309:921-926(1995).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|- , Lys-|- , but  
 CC with more restricted specificity than trypsin.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY. TRYPSIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D31789; BAA06598.1; -.  
 CC HSSP: P20231; IAAO.  
 CC MEROPS: S01.143; -.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; Trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRYP-SPC.1.  
 DR PROSITE: PS50240; TRYPSIN\_DOM.1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS.1.  
 DR PROSITE: PS00135; TRYPSIN\_SER.1.  
 KW Hydrolyase; Serine protease; Signal; Glycoprotein.  
 FT SIGNAL 1 25  
 FT CHAIN 26 270  
 FT ACT\_SITE 69 69 MAST CELL TRYPTASE.  
 FT ACT\_SITE 116 116 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 219 219 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 54 70 BY SIMILARITY.  
 FT DISULFID 150 225 BY SIMILARITY.  
 FT DISULFID 183 206 BY SIMILARITY.  
 FT DISULFID 215 243 BY SIMILARITY.  
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 270 AA; 30166 MW; 1BE102DB86943401 CRC64;

RT Match 37.8%; Score 594; DB 1; Length 270;  
 RT Local Similarity 44.0%; Pred. No. 8e-47;  
 RT Chies 122; Conservative 48; Mismatches 87; Indels 20; Gaps 7;

QY 6 AVPLLLICGSSORAKATACGRPMILRMVGGDDTQEGEMPMQVSIORNGS--HFCGG 62  
 DB 8 ALPLSLI---HRSPLCQEWG-----IYGGQELAPGNKMPQVSLRANETVRRHRCGG 56  
 QY 63 SLIDEMWLTAAHCFRNT-SETSLXVYLGAQOLVORGPAAHAKRVQVESNPILYQSTAS 121  
 DB 57 SLIRPQWLTAAHCFRNT-SETSLXVYLGAQOLVORGPAAHAKRVQVESNPILYQSTAS 113  
 QY 122 SADVALVELAPVFTYIILPVCIPDSVIFETGMCNMCVMTGMSPSDEDLPEPRLIOXL 181  
 DB 114 GADTALLELKRPVNISSVHVSILPRASETPPSTLCVNTGMDINDVSLPPEPLKEV 173  
 QY 182 AVPLIDPKKLLYSKDTFERYQPKTKNDMLCAGFEFGKRDACKDGSGLVCLVQSQW 241  
 DB 174 QVPVENVQLDCLDKHKVGYGDNHIVRDMCLAG-NEG-HDSQGDGSGPLVCKVNGTW 231  
 QY 242 LQAGVISMGEGCARONRGVYIRVTAHNNMTHRIIPK 278  
 DB 232 LQAGVISMGEGCALPNRPGITTRVYLLDMTHRIIPK 268

RESULT 10  
 MCT7\_RAT

ID MCT7\_RAT STANDARD; PRT; 273 AA.  
 AC P27435; P27436;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-NOV-1992 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mast cell protease 7 precursor (EC 3.4.21.59) (RMCP-7) (Trypsinase,  
 DE skin).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SELAGUE-DAWLEY;  
 RX MEDLINE=97149430; PubMed=8996238;  
 RA Lutzelschwaeb C., Pejler G., Aveskog H., Hellman L.;  
 RT "Secretory granule proteases in rat mast cells. Cloning of 10  
 RT different serine proteases and a carboxypeptidase A from various rat  
 RT mast cell populations." J.  
 RT J. Exp. Med. 185:13-29(1997).  
 RN [2]  
 RP SEQUENCE OF 29-53.  
 RC STRAIN=SELAGUE-DAWLEY; TISSUE=SKIN;  
 RX MEDLINE=91242400; PubMed=2036367;  
 RA Braganza V.J., Simmons W.H.;  
 RT "Tryptase from rat skin: purification and properties." J.  
 RT Biochemistry 30:4997-5007(1991).  
 RN [3]  
 RP SEQUENCE OF 29-51.  
 RC TISSUE=Breast carcinoma;  
 RX MEDLINE=9221826; PubMed=1314562;  
 RA Eto I., Grubbs C.J.;  
 RT "Separation, purification and N-terminal sequence analysis of a novel  
 RT leupelin-sensitive serine endopeptidase present in chemically  
 RT induced rat mammary tumour." J.  
 RT Biochem. J. 283:209-216(1992).  
 RL FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
 CC CELLS, AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION  
 CC RESPONSE OF THIS CELL TYPE.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|- , Lys-|- , but  
 CC with more restricted specificity than trypsin.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON  
 CC MAST CELL ACTIVATION.  
 CC -1- TISSUE SPECIFICITY: MAST CELLS.  
 CC -1- PTM: GLYCOSYLATED (PROBABLY).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY. TRYPSIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U07910; AAB8263.1; -.  
 CC PIR: A23698; A23698.  
 CC PIR: S21275; S21275.  
 DR HSSP: P20231; IAAO.  
 DR MEROPS: S01.026; -.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; Trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRYP-SPC.1.  
 DR PROSITE: PS50240; TRYPSIN\_DOM.1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS.1.  
 DR PROSITE: PS00135; TRYPSIN\_SER.1.  
 KW Hydrolyase; Serine protease; Glycoprotein; Zymogen; Signal;  
 KW Multigene family.

FT	SIGNAL	1	18	POTENTIAL.
FT	PROPEP	19	28	ACTIVATION PEPTIDE.
FT	CHAIN	29	273	MAST CELL PROTEINASE 7.
FT	ACT_SITE	72	72	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	119	119	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	222	222	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	DISULFID	57	73	BY SIMILARITY.
FT	DISULFID	153	228	BY SIMILARITY.
FT	DISULFID	186	209	BY SIMILARITY.
FT	DISULFID	218	246	BY SIMILARITY.
FT	CARBOHYD	49	49	N-LINKED (GLCNAC. . .) (PROBABLE).
FT	CONFLICT	42	42	W -> V (IN REF. 3).
FT	CONFLICT	49	51	NDT -> MLP (IN REF. 3).
SO	SEQUENCE	273 AA:	30400 MW:	65A5EDAD279FB284 CRC64:

Query Match 37.6%; Score 590.5; DB 1; Length 273;  
Best Local Similarity 45.6%; Pred. No. 1.7e-46;  
Matches 125; Conservative 43; Mismatches 95; Indels 11; Gaps 6;

9	LLLLLCGSRARAKATACGPRRLNRMVGGQDQEGEMFWQVSIQRNGS---HFCGGSLL	65
5	LLLTPLLSLVHAAPSLAMP--EGIVGGQEGASGNKMPQVSLRVNDTYMHHFCGGSLL	62
QY	AEQWVLTAAHCF-RMTSETSLYQVILGARQLVQPGPHAMYARVROVESNPLYOGTASSAD	124
Db	HPQWVLTAAHCF-RMTSETSLYQVILGARQLVQPGPHAMYARVROVESNPLYOGTASSAD	119
QY	VALVELAPVPTNYILVCLPDPSTVETGMNCWVTGWSPESEDLLEPPRILOKLAVP	184
Db	IALKLTNPVNTISNHTVSLPPASETFPSGTLCWVTGWSPESEDLLEPPRILOKLAVP	179
QY	IIDPRKCNLYSKDTERGYQPKTIKNMLCAFEEGKKDKGDSGGPLVCLVGSMTLQA	244
Db	IENRRLCDLKHKGLNTGDNVHVRDMLCAG-NEG-HDSCGGDSGGPLVCKVEDTWTLOA	237
QY	GVISMGECAQONRPGYIRVTAHNMVIRRIIPK	278
Db	GVISMGECAQONRPGYIRVTAHNMVIRRIIPK	271

Search completed: August 13, 2002, 09:00:50  
Job time: 227 sec

OM of: US-10-041-006-7 to: GenEmbl:★ out\_format : pfs  
Date: Aug 13, 2002 5:32 PM  
About: Results were produced by the GenCore software, version 4.5  
Copyright (c) 1993-2000 CompuGen Ltd.

```
-MODEL=Itrame_rpn_model -DEV=x1h
-Q/-gen2.1/USPTO_spool/US10041006/runnat.13082002_083952.11249/app_query.fasta.1.350
-DB=GenEmb1 -QEMT=fastcap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GCAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosome2
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR.SCORE=pct
-THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFM=pfs
-Next -HEAPSIZ=500 -MINLEN=0 -MAXLEN=20000000000
-US10041006 -ECGN1_1_4097 -NCPU=6 -ICPU3=3 -LONGLOG
TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1
```

```
Query length: 290
Database: GenEmbl
Database sequences: 1797656
Database length: 187333701
Search time (sec): 1855.110000
```

score\_list:

[illegible]

```

seq_name: gb_pf_HSA306593
seq_documentation_block:
LOCUS HSA306593 1013 bp mRNA linear P11-MAR-2001
DEFINITION Homo sapiens mRNA for marapsin (MPN gene).
ACCESSION AJ306593
VERSION AJ306593.1 GI:13516325
KEYWORDS marapsin; MPN gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Fortunato,M., Dando,P.M., Rawlings,N.D. and Barrett,A.J.
TITLE Cloning, sequencing and expression of marapsin, a human serine
protease
JOURNAL Unpublished
REFERENCE
AUTHORS Fortunato,M.
TITLE Direct Submission
JOURNAL Submitted(29-MAR-2001) Fortunato M., MRC Molecular Enzymology
Laboratory, Babraham Institute, Babraham Hall, Babraham, Cambridge,
CBR 4AT, UNITED KINGDOM
COMMENT Related entry AM170323
FEATURES
Source Location/Organism
gene 1..1013 /organism="Homo sapiens"
cds /db_xref="taxon:9606"
20..892 /clone="IMAGE 2698030"
20..892 /gene="MPN"
20..892 /gene="MPN"
/function="putative trypsin-like activity"
/codon_start=1
/product="marapsin"
/protein_id="CAC35467.1"
/db_xref="GI:13516326"
/translaton="MRRRAVPIPLLLICFSSQAKAATAAGRPRLNRWVGDDTDEG
EMPMQYSIQRNSHFCGGSLIAQMVLTAHCERNTSLYVLGARQLVPGFHA
MYRVKQVESNPPLYOGTASSADVALVELEAPVFPTNYILPVCLPDSYIFETGMNCMV
EGGSSSEEDLLEPRILQKLAVPIIDTPCNIILYSKDFEQPTIKNDIMCAFE
EGKKDKCKDGGSPVLCLVGOISWLQAGVISIWSGCARQNRPYIVRTAHNNIHRII
PKLPQPARLGCGK"
BASE COUNT 192 a 334 c 321 g 166 t
ORIGIN
alignment_scores:
Quality: 1571.00 Length: 290
Ratio: 5.417 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-10-Q41-006-7 x HSA306593 ..
Align seg 1/1 to: HSA306593 from: 1 to: 1013
1 MetatgargtgcctgaalavalaProleuleuleuleuleuCySpheglyse 17
|||||
20 ATGAGCGCGCGCGCGCGCTCCGTCTGCTGCCTGTTGTC 69
|||||
17 TGLnrgAlaLySaLaThrAlaCySgLYArGPraGrMetLeuAsna 34
|||||
70 TCAGAAGGCCAAGGACGACACAGCCTGTGTGTGCCCCGAGATGCTGAACC 119
|||||
34 rgmetValcLYcLYcLnAspThrGlnGlnGlyGluTrpProTyrGlnVal 50

```



```

797 TTTGGCTACCAACCCAAAACCATCAGAAATGACATGCTGTGGCCGCGCAT 846
|||||
217 PheGluGluGlyLysLysAspAlaCysLys.GlyAspSerGlyGlyProl 233
|||||
847 TTCGAGAGAGGCGCAAGAGAGATGCTCCAGATGGG.....GGAGCCA 887
|||||
233 euValCysLeuValAlaGlyInSerTriPLeuGlnAlaGlyVal...IleSer 248
:::|||||:::
888 TGGGGTGCTTCCATGGGGATTCCACTTCCAAACCCGACAGGCCAGGCCAAGG 937
|||||
249 TPrgIyGluGly.....CysAlaArgGlnAsnArgProG1 260
|||||
938 TGGGGTTTGAAGCCCTGTCCATGGTTATGCTTCAACAGAAACCGGCCAG 987
|||||
260 yValTyrlIeArgValThralHisHisAsnTPriLeHisArgIleIleP 277
:::|||||:::
988 TGCTTTAAATCCGCTGTGACGCGCCACAAACAAATGGATTCATCGGATATGCG 1037
|||||
277 roLysLeuGlnPheGlnProAlaArgLeuGlyGlyLys 289
|||||
1038 CCNAACCTGAG.TTCCAGCCACCAAGCAAGTTGGCGCCGACG 1074
|||||
seq_name: gb_pat:E31782

```

seq_documentation_block:					
LOCUS	E31782	1109 bp	DNA	linear	PAT 07-FEB-2001
DEFINITION	Novel compound.				
ACCESSION	E31782				
VERSION	E31782.1	GI:13025811			
KEYWORDS	JP 199008992-A/1.				
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 1109)				
AUTHORS	Nicola,A.B.H.H. and Kurinkenbiado,C.D.S.S.				
TITLE	Novel compound				
JOURNAL	Patent: JP 199008992-A 1 13-APR-1999;				
	SMITHKLINE BEECHAM CORP PUBLIC LTD CO				
COMMENT	OS unidentified				

PR 09-JUN-1998 JP 1998I1994Z5  
PR 10-JUN-1997 GB 97I2008L5.17-OCT-1997 GB 97308295-1, PR  
20-FEB-1998 GB 9803650.2  
PI NICOLA ANN BAUESU, HELEN ELIZABETH KURINKENNAIO, PI  
CHRISTOPHER DONALD SAZAN  
PC C12N5/09,A6IK31/7/0,A6IK35/7/6,A6IK38/00,A6IK38/00  
PC A6IK39/395.  
PC A6IK39/395,A6IK48/00,C07K16/40,C12N5/10,C12P21/02,C12P21/08  
PC C12O1/68,  
PC C12N5/00,A6IK37/02,A6IK37/02,A6IK37/02,C12N5/00 CC  
Strandedness: Single;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT source 1.I109  
'organization='Unidentified'.

FEATURES	Location/Qualifiers
SOURCE	1. .1109
	/organism="unidentified"
BASE COUNT	/db_xref="taxon:32644"
ORIGIN	229 a 344 c 344 g 192 t

alignment_scores:		
Quality:	1228.50	297
Ratio:	4.601	Gaps: 5
Percent Similarity:	89.899	Percent Identity: 83.502

alignment\_block:  
US-10-041-006-7 x E31782

Align seg 1/1 to: E31782 from: 1 to: 1109

[illegible]

```
/db_xref="taxon:32644"
```

BASE COUNT	224 a	327 c	338 g	191 t	29 others
------------	-------	-------	-------	-------	-----------





```

DEFINITION Homo sapiens chromosome 16 clone CTD-259JF12, complete sequence.
ACCESSION AC093517
VERSION AC093517.2 GI:15808524
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 160601)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 16
REFERENCE
AUTHORS 2 (bases 1 to 160601)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submmission
REFERENCE
AUTHORS 3 (bases 1 to 160601)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submmission
REFERENCE
AUTHORS 3 (bases 1 to 160601)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submmission
COMMENT
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES
source
Location/Qualifiers
1..160601
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="CTD-259JF12"
BASE COUNT 37601 a 43383 c 43006 g 36611 t
ORIGIN
alignment_scores:
Quality: 997.00 Length: 1090
Ratio: 3.734 Gaps: 4
Percent Similarity: 24.495 Percent Identity: 24.404
alignment_block:
US-10-041-006-7 x AC093517/rev ..
Align seg 1/1 to reverse of: AC093517 from: 1 to: 160601
22 AAlaAlaThrAlaCysGlyArpArGmetLeuAsnArgMetValGly 38
|||||
100378 GCTGCCCCACCTGTGTCGCCCCACAGAGTCGAAACCGAATGTTGGCGG 100379
38 YGlnAspThrGlnGluGlyIuTrpProTrpGlnValSerIleGlnArg 55
|||||
100328 GCAGGACACCCAGAGAGGCGAGTGGCCCTGGCAAGTCACATCCAGCC 100329
55 nGlySerHisPheCysGlyGlySerLeuIleAlaGluGlnTrpValLeu 71
|||||

```

100278 ACGGAGCCACTTCTGGGGGAGAGCCTCATCGGGAGAGTGGTCTG 100229  
72 ThrAlaAlaHisCysPhe..... 77  
100228 ACGGCTGCGACTGCTTCCCAAGTAGTCCGCCGCCCTGCCGCCGCC 100179  
77 ..... 77  
100178 CATAGCCGTGACAGCGCCCGCGGCGACGGGTTCAGCACCGTGGACAGC 100129  
77 ..... 77  
100128 GCCCGCGCGCCAAATCTGGGGGTGACTTCCTGGGGGCTCTGTGCA 100079  
77 ..... 77  
100078 GCCCCTCCACCCAGATGCTTCCCTTAGTCCAACTCCAGGGCTAATTC 100029  
77 ..... 77  
10028 CAGTTGCAACGGCTGCTCCCGCCCGGGAGGTGCTCGCACCGCCCCC 99979  
77 ..... 77  
99978 GACCCCTCATCCCTTCACACCACTACCCACTCCCTGTGGTCCCTGC 99929  
77 ..... 77  
99928 AGAAGCGCCGCGAGGCTGCCCCACCGCCCTCTGCGCTTCCCA 99879  
77 ..... 77  
99878 TCCCGCACACACTCAGCTCCAGGACACTTTCGGGAGAACTGCT 99829  
77 ..... 77  
99828 CACAAAGCCCAAGACAGACAGAAAGCCCTTCCTCCCTCACCCACT 99779  
77 ..... 77  
99778 GAACCAACCCAGAAAGCCCTGAGCAGAGGCCACCAACCCCTCTGC 99729  
77 ..... 77  
99728 CATGATGACACCACTGCTGCCACACTTCCGGGTGCCAGGCCCTC 99679  
77 ..... 77  
99678 ACCTCACAACCTCAACACCCGAGCTTAATTAATTAACCCACATCTTT 99629  
77 ..... 77  
99628 TTCTTTTCTCTCTGATCTTTAAAGAAATATCATGACAAAAAAC 99579  
77 ..... 77  
99578 CCACATCTTAATTCAGATCTCAAGCGCCAGCAGCGTGGCTCACACCCG 99529  
77 ..... 77  
99528 TAATCCAGCACTTGGAGGCGCAAGCGGCGAGATCAGTTGAGCCAGG 99479  
77 ..... 77  
99478 AGTTAAGACCAAGCCGGGCAACACAGACACCTGTCTACTAAAAA 99429  
77 ..... 77  
99428 TAAAAAAATTAATAAACAATAAACAATAAACAATAAACCAGGTG 99379  
77 ..... 77  
99378 CAGTGTGCTGTGTGCTCCCAAGTACTCAGGAAGCTGAGCGAGAGATGG 99329

77 ..... 77  
99328 CTTGAGCCAGAGAGTTGGAGGCTGCAGTAGTACAGTTGACTGCTGCA 99279  
77 ..... 77  
99278 CTCACGCTGGGTGACAGAAATGAGAACTGTCTCTAGGGGAAAAAAA 99229  
77 ..... 77  
99228 ATCGAAACTCGGGCTGAGAGAGGCGAGTCACTGCCAGAGCCCGCAG 99179  
77 ..... 77  
99178 GTGGGGCTGGCCTGCTGCACCTGGAGGGTGGTCTCTCCGAGCCAGCT 99129  
77 ..... 77  
99128 TTCAGTAGGCACTGTACTTCCACACAGACAGGTAGAAAGCTGAT 99079  
77 ..... 77  
99078 CACAGAAACTTGTTAAGTACTTGGCCAAAGGCCACACAGCGGCCAGGG 99029  
77 ..... 77  
99028 TGGAGTGGAAATCTGAGTGGGGGCTGGCTTCAATGTCAACAGCAT 98979  
77 ..... 77  
98978 CACCGCTGCACCTGCTGCTCTAGAGATGACTCTCGAAGACATTTGACAG 98929  
77 ..... 77  
98928 AGAGAGGTGTGGGCAAGGTGTGCCAAGAGACCCACCCCTCCCG 98879  
78 ..... Ar 78  
98878 GCCTGCCCCACACAGAGGGCCGCTCACGCCCTCTGTTTCGCG 98829  
78 g.AsnThrSerGluThrSerLeuTyrGlnValLeuLeuGlnValArgGln 94  
98828 CCAGCACTCTGAGAGCTCCCTGTACAGAGTCCCTGCGGGGCGAAGCAG 98779  
95 LeuValGlnProGlyProHisAlaMetTyrAlaArgValArgGlnVal 111  
98778 CTAGTGCAGCGGGGACACACGCTATGTATGCCGGGTGAGCGAGTGA 98729  
111 userAsnProLeuTyrGlnGlyThrAlaSerSerAlaAspValAlaLeu 128  
98728 GAGCAACCCCTGTACAGGCGACGGCTCCACGCTGACGTGGCTGG 98679  
128 AlGluLeuGlnAlaProValProPheThrAsnTyrIleLeuProValCys 144  
98678 TGGAGCTGGAGGACCAAGTGCCTTCAACCAATTAATCCCTCCGCTGC 98629  
145 LeuProAspProSerValIlePheGluThrGlyMetAsnCysTyrValIb 161  
98628 CTGCTTACCCCTCGGTGATCTTTGAGACGGGCATGACTGCTGGTCA 98579  
161 rGlyTyrGlySerProSerGlu..... 169  
98578 TGGCTGGGCGAGCCCGCAGTGAAGGTAAAGGGACAGGCTGGAAAG 98529  
169 ..... 169  
98528 ATGGGGATATGCTCAAGAAGGCCAGAGCGCTTGGGACCACTGCA 98479  
169 ..... 169  
98478 ACCCAATCTCTCTGTTGCTAATTAATTAATACATACACAGCTCTGTT 98429

```

169 ..... 169
98428 ACACAGTGGGGCCGAGGGAGCAGACAGACCAGAAAGAGGGGG 98379
169 ..... 169
98378 TGAAGGAGAGGGGGACAGAAAGGCCAGCTGCTGCGAAGAGGG 98329
169 ..... 169
98328 AGGATGCTGCCCCACAGTCCCTAAGCCAGATGTTCTAAACATGAG 98279
169 ..... 169
98278 AAACGACGAGAGATGAGTGTGTTCGGGGCCGCTGGTCTTGAT 98229
170 ..... AspleuLeu 172
98228 GAGGAGTCCGTTGAGCCCTGGCTGTCGCCCTCTCCAGCCTCTG 98179
173 ProGluProArgIleLeuGlnLysLeuAlaValProIleIleAspThrP 189
98178 CCGGACCGCGGATCCGCGAAGAACTCGTGTCCCATCATGACACACC 98129
189 OlyCysAsnLeuLeuTyrSerLysAspThrGluPheGlyTyrGlnProL 206
98128 CAAGTGCACCTGCTCTACAGCAAGACACCGAGTTGGCTACCAACCA 98079
206 YstHriIleLysAsnAspMetLeuCysAlaGlyPheGluGlyLysLys 222
98078 AAACCATCAAGATGACATGCTGTGCGCGGCTCGAGGAGGGCAAG 98029
223 AspAlaCysLys ..... 226
98028 GATGCTGCAGAGGTGGAGCAATGTGTGTCATGGGACTGATCCCA 97979
226 ..... 226
97978 CCGCCAGGCCGAGGCGAGGTGGATATGCCCCGCTCATGGGCTATGC 97929
226 ..... 226
97928 ATTCAACAGCTGAGCAGCTTTTGAAGAGCAGATTCCAGGGCCGG 97879
226 ..... 226
97878 TGTGTGGCTCAGCCCTGTAATCCAGCACTTGGAGGCCAAGAGGCT 97829
226 ..... 226
28 GGATCAAGAGTCAAGATGAGAGCCATCTGGCTAACACAGTGAACCC 97779
226 ..... 226
97778 CCATCTACTATAAATAACAAAAATTAGCCAGGCGTGTGTGGGGCC 97729
226 ..... 226
97728 TGTATGCCAGCTACTTGGAGGCTGAGCAGAGAAATGGCCGAACTG 97679
226 ..... 226
97678 GAGAGGGGAGCTTGCAGTGAGCCGAGATGCCGCACTGCCAGCCTG 97629
226 ..... 226
97628 GGCACAGAGCAAGACTCCGCTCAATAAAAAAAGAAAAAGCAGA 97579
226 ..... 226
97578 TTCCTGGACCCCAACCCAGAGATCAGATGATGGGTGGGGTGGGG 97529
226 ..... 226

```

```

97528 CACAGAAACTGATTTTCAACACACACACACACACACACACACAC 97479
226 ..... 226
97478 ACACTAAGAAATAGCACTTAGACATCGGACTGGAGGGAGTGGAG 97429
226 ..... 226
97428 CCGGCGAGAGAGCTCTGGGATGAGAGAGAAATCCATTTCGAGCTAC 97379
226 ..... 226
97378 ACACCTCTGTGCGGCTGCAACTGAGTCTTGGGAATCAAGTGGGCTC 97329
227 ..... GlyAspSerGlyGlyProLeuVal 234
97328 TGACGGTACTGTGCTGCCCGCCGAGGGGAGCTCGGGCGCCCTGGTG 97279
235 CysLeuValGlyGlnSerTrpLeuGlnAlaGlyValIleSerTrpGly 251
97278 TGCCCTGCGGCTCACTGCTGGCTGAGGCGGGGGTGAATCACTGGGCTGA 97229
251 uGlyCysAlaArgGlnAsnArgProGlyValTyrIleArgValThrAla 268
97228 GGGCTGTCCCGCCAGAACCGCCAGGTGTCTACATCGTGTCAACGCC 97179
268 IAspAsnTrpIleHisArgIleLeuProLysLeuGlnPheGlnProAla 284
97178 ACCAACACTGATGATCCATGATCCCAAACTGACATTCACAGCAGCG 97129
285 ArgLeuGlyGlyGlnLys 290
97128 AGGTGGCGGCCAGAAAG 97111
seq_name: gb_hhg:AC092117
seq_documentation_block:
LOCUS AC092117 208529 bp DNA linear HTG 23-AUG-2001
DEFINITION Homo sapiens chromosome 16 clone CTD-2270P14, WORKING DRAFT
SEQUENCE. 2 ordered pieces.
ACCESSION AC092117
VERSION AC092117.2 GI:15281234
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 208529)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
JOURNAL 2 (bases 1 to 208529)
REFERENCE DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Submitted (22-JUN-2001) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Aug 23, 2001 this sequence version replaced gi:14522963.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have

\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 154083: contig of 154083 bp in length  
\* 154084 154183: gap of unknown length  
\* 154184 208529: contig of 54346 bp in length.

FEATURES  
source 1. 208529  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="CTD-2270P14"

BASE COUNT 48217 a 56176 c 56702 g 47334 t 100 others  
ORIGIN

ment\_scores:  
Quality: 997.00 Length: 1090  
Ratio: 3.734 Gaps: 4  
Percent Similarity: 24.495 Percent Identity: 24.404

alignment\_block:  
US-10-041-006-7 x AC092117 ..

Align seg 1/1 to: AC092117 from: 1 to: 208529

```
22 ALAAlaThrAlaCysGlyArgProArgMetLeuAsnArgMetValGlyI 38
|||||
105010 GCTCCCCAGCCTGTGGTGGCCCGGAGATGCTGAACCGAATGGGGGCGG 105059
38 YGlasPthrcIngluglyuTrpProTrpGlnValSerIleGlnArgA 55
|||||
105060 GCAAGACACGAGGAGGCGGAGTGGCCCTGGCAAGTCAGACGATCCAGCGCA 105109
55 snGlySerHisPheCysGlyGlySerLeuIleAlaGluIntPrValIleu 71
|||||
105110 ACGGAGGCACTTCTGGGGGGGAGCCTCATCGGAGCAAGTGGGCTCTG 105159
72 ThrAlaAlaHisCysPhe..... 77
|||||
105160 ACGGCTGGCAGCTGCTTCCGCAAGTGAATCCGCCGCCCTGCCCCGCC 105209
77 ..... 77
105210 CATAGCCTGACAGCGCCCGCGCGACCGGTTCAAGACCGGTGAGACA 105259
77 ..... 77
05260 GCCGCGCGCGCAATCCTGGGGGTGAACCTCCCTGGGGGCTCTGTTCCA 105309
77 ..... 77
105310 GCCCCTCCACCCAGATGCTTCCCTTAGTCCAACTCCAGGGCTAATCTC 105359
77 ..... 77
105360 CAGTTGCAACCGCTGCTCCGCCCGGAGGAGTGCCTGCACCGCCCCC 105409
77 ..... 77
105410 GACCCCTCATCCCTCCACCACTCAACCACTCCCTGTGGTCCCTGC 105459
77 ..... 77
105460 AGAAGCGCGCGGAGGCTGTGCCACCGGCGCTCTGGCTTTCCCA 105509
77 ..... 77
105510 TCCGCGACACACTGAGCTCCAAGACACTTCCCGGAGGAACTGCT 105559
77 ..... 77
```

```
105560 CACAAAGCCCAAGAGACAGACAGAAAGGCGCTTCTCCCTCACCACCT 105609
77 ..... 77
105610 GAACCAACCCCAAGAAAGCCCTGAGAGAGCCAGCCACCACCGCTCTGC 105659
77 ..... 77
105660 CATGTATGAAACACCTGTTCCACACACTTCCGGGTGTCCAGGCCCTTC 105709
77 ..... 77
105710 ACCTCAGACCTCAACACGCGAGCTCTAATTATTTAAACCCACATCTTT 105759
77 ..... 77
105760 TTTCTTTTCTTCTTGTGATCTTTAAAGAAATATCATGACAAAAAAC 105809
77 ..... 77
105810 CCACATCTTAATTCAGATATCTACGCGCCAGGACGGTGGCTCACACCG 105859
77 ..... 77
105860 TAATCCAGCACTTTGGAGGCCAAGCGGCGAGATCATGATGACCCAGG 105909
77 ..... 77
105910 AGTTCAAGACCAACCGCGGCGACACAGAACCTGTCTCTACTTAAAAA 105959
77 ..... 77
105960 TAAAAAAATTTAAAAACAAAAAACAAAAAACAAAAAACCCCAAGTG 106009
77 ..... 77
106010 CAGTGTGCTGTGTGTCCTCAAGTACTCAGAAAGTGTGAGGAGAGTGG 106059
77 ..... 77
106060 CTTGAGCCCAAGAGTGTGAGGCTGCAAGTACTACAGTTGCACTGCTGA 106109
77 ..... 77
106110 CTCAGGCTGGGTGACAGATGGAATCTGTCTTAAGGGGAAAAAAA 106159
77 ..... 77
106160 ATCGAAATCGGGGTGAGAGAGGCGAGTCACTTCAGAGCGCCGCGAG 106209
77 ..... 77
106210 GTGGGCTGGCCCTGCTGCACTGAGGGTGTCTCTCCAGCCAGCT 106259
77 ..... 77
106260 TTGAGTAGGCACTGTGACTGTCCCAACAGCACAGGTGAGAAAGTGCAT 106309
77 ..... 77
106310 CACAGAACTGTTAAGTACTTGGCCAAAGGCCACAGACGCGCCAGGG 106359
77 ..... 77
106360 TGAAGTGAATATGAGTGGCGGCTGTGGCTTCAATGTCAACAGCAT 106409
77 ..... 77
106410 CACCGCTGACACTGGCTGTCTAGAGATGACTCTGGAAGACATTTGCA 106459
77 ..... 77
106460 AGAAGGTGTGGGCGACGGGTGTGGCCAGAGAGCCCGACCGCTCCCG 106509
```

78 .....Ar 78  
106510 GCCTGTCACACAGAGGCGCCCTCAGCCCTCTCTGTTTCCCG 106559  
78 g.AsnThrSerGluThrSerLeuTyrGlnValLeuLeuGlyAlaArgGln 94  
106560 CCAGCACTCTGAGAGCTCCCTGTACCAAGTCTGCTGGGCGCAAGCAG 106609  
95 LeuValGlnProGlyProHisAlaMetTyrAlaArgValArgGlnValG 111  
106610 CTAGTGAGCGGGACACACAGCTATGATGCCCCGGGTGAGCGAGTGA 106659  
111 userAsnProLeuTyrGlnGlyThrAlaSerSerAlaAspValAlaLeu 128  
106660 GACGACCCCTGTACAGGGCAGCGGCTCCAGCGCTGAGCTGGCCCTG 106709  
128 a.lGlLeuGlnAlaProValProPheThrAsnTyrIleLeuProValCys 144  
106710 TGGAGCTGGAGGCGACAGTGCCTTCCACCAATTCCTCCCGCTGTGC 106759  
145 LeuProAspProSerValIlePheGluThrGlyMetAsnCysTrpValTh 161  
106760 CTGCTGACCCCTCGGATCTTTGAGAGGGGATGATGACTGCTGGTCA 106809  
161 r.glyTyrGlySerProSerGluGlu..... 169  
106810 TGGCTGGGGCAGGCCCGAGTGAAGAGTAAAGGGACAGGGCTGGAAAGA 106859  
169 ..... 169  
106860 ATGGGGGATATGCTCTCAGAAAGGCCAGAGCGGCTTGGGAGCACCTGTA 106909  
169 ..... 169  
106910 ACCCACAATCTTCTGTGTCTATATTAATACATACACAGGCTCTGTTC 106959  
169 ..... 169  
106960 ACACAGGTGGGCGCGAGGGAGCAGCAGCAGACCAGAAAGAAAGGGG 107009  
169 ..... 169  
107010 TGAAGGAGAGGGGGACAGAAAGGCCAGCTGCTGCTGGAAGAGAG 107059  
169 ..... 169  
107060 AGGATGTCGCCCCCAGTGCCTTAAGCCAGAGATGCTTAACATGAGG 107109  
69 ..... 169  
110 AAACGACAGGCAAGATGAGTGGTGTCCGGGGCGTGGGTCTTGAT 107159  
170 .....AspLeuLeu 172  
107160 GAGGAAGTCCTGTGAGCCCTGGGCTGTTCGCCCTCTCCAGACTCTCG 107209  
173 ProGluProArgIleLeuGlnLysLeuAlaValProIleIleAspThrPr 189  
107210 CCGAAGCCGGGATCTCTGAGAACTGCTGTGCCATATGACACAGC 107259  
189 oLysCysAsnLeuLeuTyrSerLysAspThrGluPheGlyTyrGlnProL 206  
107260 CAAGTGCAACCTGCTCTACAGCAAGACACCGAGTTGGCTACCAACCA 107309  
206 ystThrIleLysAsnAspMetLeuCysAlaGlyPheGlnGlnLysLys 222  
107310 AAACCTCTCAAGATGACATGCTGTGGCGGCTCTGAGGAGGGCAAGAG 107359  
223 AspAlaCysLys..... 226  
107360 GATGCTCGCAAGGTGGAGCAATGTGTCTCCATGGGAGCTCAGTCCCA 107409

226 ..... 226  
107410 CCGCAGGCCCCAGGGCAGGTGGATCATGCCCGCTCCATGGGCTATGC 107459  
226 ..... 226  
107460 ATTCACACAGCTGACACAGCTTGTTTGAAAGGAGATTCCAGGGCGGG 107509  
226 ..... 226  
107510 TGTGTGCTCAGCGCTGTAAATCCAGCACTTTGGAGGCCAAGAGAGGT 107559  
226 ..... 226  
107560 GSATPACAAGGTCAAGATGAGACATCTCTGGCTAACACAGTAAACC 107609  
226 ..... 226  
107610 CCATCTCTACTAAAAATACAAAAAATTAGCCAGGCGTGTGTGGCGCC 107659  
226 ..... 226  
107660 TGTACTCCACCTACTTGGAGGCTGAGGCGAGAGAAATGGCGGAACTG 107709  
226 ..... 226  
107710 GAGGCGGAGGCTTGACGTGAGCCGAGATCGGCGCACTGCACCTCCAGCTG 107759  
226 ..... 226  
107760 GGCCACAGAGACACTCCGTCTCATATMAAAAAAAAAAGAAAGCAGA 107809  
226 ..... 226  
107810 TTCTGAGCCCGCCACCCAGAAAGATCAGAGTCAATGGCTGCGGCTGGCG 107859  
226 ..... 226  
107860 CACAAGAACTGCAATTTTCAACACACACACACACCTAGTTGATCTGGG 107909  
226 ..... 226  
107910 ACACTAAGAAATAGGCACTTTAGCAATCGGACTGGAGGGGATGGAG 107959  
226 ..... 226  
107960 CCGGCGAGAGGCTTGAGATGAGAGAAAGAAATCCATTGGGAGCTAC 108009  
226 ..... 226  
108010 ACACCTCTGTGCGGCTGCAACTGAGTCTGGGAACTCAAGTGGGCTC 108059  
227 .....GlyAspSerGlyGlyProLeuVal 234  
108060 TGACGTGACTCTGCTGCTCCCGGAGGGCGACTGCGGGCGCCCTGTGTG 108109  
235 CysLeuValGlyGlnSerTyrLeuGlnAlaGlyValIleSerTrpGly 251  
108110 TGCTCTGCGGTACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 108159  
251 uGlyCysAlaArgGlnAsnArgProGlyValTyrIleArgValThrAlaH 268  
108160 GGGCTGTGCCCGCAGAACCGCCAGGTGTCTACATCGTGTACACGCCC 108209  
268 sHisAsnTrpIleHisArgIleIleProLysLeuGlnPheGlnProAla 284  
108210 ACCAAGCTGATCATCGGATCATCCCAAACTGACGTTCCAGCAGCG 108259  
285 ArgLeuGlyGlyGlnLys 290  
108260 AGTTGGGCGGCGCAGAG 108277  
seq\_name: gb\_pr.AC004036

```

seq documentation block:
LOCUS      AC004036          40179 bp    DNA        linear   PRI 27-JAN-1998
DEFINITION Homo sapiens chromosome 16, cosmid clone 400DI (LANL), complete
sequence.
ACCESSION  AC004036
VERSION    AC004036.1  GI:2811104
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 40179)
            Rieke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
            Goodwin,L., Bryant,J., Tesmer,J., Melnick,L., Longmire,J.,
            White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
            Misra,M. and Deaven,L.
            Sequencing of Human Chromosome 16p13.3
            Unpublished
            2 (bases 1 to 40179)
            Rieke,D.O. and Wagner,R.P.
            Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) system
            Unpublished
            3 (bases 1 to 40179)
            Rieke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
            Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
            Goodwin,L., Bryant,J., Tesmer,J., Melnick,L., Longmire,J.,
            White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
            Misra,M. and Deaven,L.
            Direct Submission
            Submitted (27-JAN-1998) Center for Human Genome Studies, DOE Joint
            Genome Institute, Los Alamos National Laboratory, MS M888, Los
            Alamos, NM 87545, USA
FEATURES
    source
        1..40179
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="16"
            /map="16p13.3"
            /clone="400DI"
            1..92
            /rpt_family="AluJo"
            complement(547..899)
            /rpt_family="Alu"
            complement(900..1124)
            /rpt_family="AluSx"
            1180..1463
            /rpt_family="Alu"
            1479..1764
            /rpt_family="Alu"
            complement(1806..2099)
            /rpt_family="Alu"
            complement(2178..2464)
            /rpt_family="Alu"
            complement(2875..3215)
            /rpt_family="Alu"
            3996..4294
            /rpt_family="Alu"
            4578..4836
            /rpt_family="Alu"
            4898..5098
            /rpt_family="Alu"
            complement(5129..5463)
            /rpt_family="AluSx"
            complement(5449..5798)
            /rpt_family="Alu"
            complement(5869..6176)
            /rpt_family="Alu"
            6439..6779
            /rpt_family="Alu"
            6800..7118
            /rpt_family="Alu"

```

```

            /rpt_family="Alu"
            complement(7280..7602)
            /rpt_family="Alu"
            7288..7307
            /note="(1)20"
            /rpt_type=tandem
            /rpt_unit="T"
            complement(7651..7977)
            /rpt_family="Alu"
            complement(7949..8271)
            /rpt_family="Alu"
            complement(8469..8662)
            /rpt_family="AluSx"
            8782..9102
            /rpt_family="AluSx"
            9089..9112
            /note="(AAT)8"
            /rpt_type=tandem
            /rpt_unit="AAT"
            9193..9329
            /rpt_family="L1"
            9260..9654
            /rpt_family="Alu"
            9801..10106
            /rpt_family="MLT1"
            10245..10303
            /rpt_family="Alu"
            complement(11162..11322)
            /rpt_family="Alu"
            11357..11733
            /rpt_family="Alu"
            complement(11734..12016)
            /rpt_family="Alu"
            complement(12160..12264)
            /rpt_family="Alu"
            12313..12638
            /rpt_family="Alu"
            complement(12840..13147)
            /rpt_family="Alu"
            complement(13345..13714)
            /rpt_family="Alu"
            14679..15069
            /rpt_family="Alu"
            16880..17196
            /rpt_family="Alu"
            17175..17194
            /note="(A)20"
            /rpt_type=tandem
            /rpt_unit="A"
            19639..19803
            /note="SP0031A Serine Protease exon"
            19727..19792
            /note="SP0031A Serine Protease exon"
            20443..20771
            /rpt_family="Alu"
            complement(20874..21003)
            /rpt_family="MIR"
            21239..21463
            /note="SP0031A Serine Protease exon"
            21252..21446
            /note="100% identity EST724219 CAMA1B cell line II"
            /db_xref="dbEST:AA321681"
            22114..22420
            /rpt_family="Alu"
            complement(22411..22493)
            /rpt_family="MERS"
            22692..22879
            /note="SP0031A"
            complement(22899..23084)
            /note="98% identity EST80292"
            /db_xref="dbEST:AA368960"
            24791..24832
            /note="(AC)21"

```





78 ..... 78  
 20970 GTGAGGTGGAATCTGATCGGGGCTGTGCTTCAATGTCAACACGCA 21019  
 78 ..... 78  
 21020 TCACCGCTGCACCTGCTGTCTTACAGATGACTCTGAAGACATTGTGCA 21069  
 78 ..... 78  
 21070 GAGAGAGGTGTGGGGCAGCGTGTGGCCAGAGAGCCCAAGCCCTCCCG 21119  
 78 ..... 78  
 21120 GGCCTGTCCCAACAGAGGGGCCCTCAGCGCCCTCTGTCTTCTCC 21169  
 79 .Asn...ThrSerGluThrSerLeuTyrGlnValLeuLeuGlyAlaArgI 94  
 170 GCCAGCACCCTCTGAGAGCGTCCCTGTACCAAGCTGTGCGGGGCAAGGCA 21219  
 94 nLeuValGlnProGlyProHisAlaMetTyrAlaArgValArgGlnValG 111  
 21220 GCTAGTGCAGCGCGGACCAACACGCTATGTATGCCGGGTGAGGCAAGTGG 21269  
 111 LuSerAsnProLeuTyrGlnGlyThrAlaSerSerAlaAspValAlaLeu 127  
 21270 AGAGCAACCCCTGTACCAAGGCGACGCGCTCCAGCGCTGACGTGGCCCTG 21319  
 128 ValGluLeuGluAlaProValProPheThrAsnTyrIleLeuProValIcy 144  
 21320 GTGGAGCTGGAGGACACAGTGCCTTACCAATTCATCCTCCCGTGTG 21369  
 144 sLeuProAspProSerValIlePheGluThrGlyMetAsnCysTrpValT 161  
 21370 CCTGCTGACCCCTGCTGTGATCTTGAACGCGCATGAACCTGTGGTCA 21419  
 161 hGlyTyrPglySerProSerGluGlu..... 169  
 21420 CTGGCTGGGGGAGCCCGAGTAGGAAGTAAGGGAGCAGGGCTGGGAAG 21469  
 169 ..... 169  
 21470 AATGGGGCATATGCTCAAGAAAGCCAGAGCGCTTGGGACCACTGTG 21519  
 169 ..... 169  
 21520 AACCCACAATCTTCTGTTGCTATATTAATACATACACAGGTCTCTGT 21569  
 169 ..... 169  
 21570 CACACAGGTGGGGCGAGGGGAGCAGACAGACCCAGAAAGAGAGGGG 21619  
 169 ..... 169  
 21620 GTGAAGGAGAGAGGGGACAGAAAGGCCCAAGCTGCTGCTGGAAGAGG 21669  
 169 ..... 169  
 21670 GAGGATGTCTGCCCAAGTAGTGCCTAAGCCAGAGATGGTCTAATCATGAG 21719  
 169 ..... 169  
 21720 GAAAGCAGAGCAGAGATGAGTGTGTCCCGGGCGCTGGGTTCTTGA 21769  
 170 ..... AspLeuLe 172  
 21770 TTAGGAAATCCCTTGAAGCCCTGGCTGTCCGCCCTCTCCAGACTCTCT 21819  
 172 uProGluProArgIleLeuGlnLysLeuAlaValProIleIleAspTrp 189  
 21820 GCCGCAACCGCGGATCTGTGCAAGAACTCGCTGTGCCCATTCATCGACACAC 21869  
 189 rGlyCysAlaAsnLeuTyrSerLysAspTrpGlnPheGlyTyrGlnPro 205

|||||  
 21870 CCAAGTCAACCTCTCTACAGCAAAAGACACGAGTGTGCTACCAACCC 21919  
 206 LysThrIleLysAsnAspMetLeuGlyPheGluIleLysLysLys 222  
 21920 AAAACCATCAAGAAATGACATGTGTGCGCGGCTTGCAGAGAGGCAAGAA 21969  
 222 sAspAlaCysLys..... 226  
 21970 GGATGCTTGCAGAGTGGGAGCATGTGTGTCCATGGGAGCTAGTCCC 22019  
 226 ..... 226  
 22020 ACCGCAAGCCCAAGGAGGAGGTGGATCATGCCCGCTCCATGGGCTATG 22069  
 226 ..... 226  
 22070 CATTCACACAGCTGAGCAGCTGTTTGTGAAAAGCAGATTCCAGGGCCGG 22119  
 226 ..... 226  
 22120 GTGTGTGGCTCAGCGCTGTATCCAGCACTTTGGGAGGCCAAGAGGG 22169  
 226 ..... 226  
 22170 TGGATCACAAGGTCAGAGATGAGACCATCTGTGCTAACACAGTGAAC 22219  
 226 ..... 226  
 22220 CCCATCTCTACTAAATAATACAAAAAATTAGCCAGCGCTGTGTGGCGC 22269  
 226 ..... 226  
 22270 CTGATGCCACGACTACTTGGGAGGCTGAGCAGAGAGAATGGCCGAACCT 22319  
 226 ..... 226  
 22320 GGGAGCGGAGCTTGCAGTAGCGGAGAGTCGCCAGTGCACCTCCAGCCT 22369  
 226 ..... 226  
 22370 GGGCCACAGACCAAGACTCCGTCTCAATAAAAAAAGAAAAAGCAG 22419  
 226 ..... 226  
 22420 ATTCTGGACCCCAAGGAGATCAGATGAGTGTGTGGGTGGG 22469  
 226 ..... 226  
 22470 GCACAGAAACTGCATTTTCAACACACACACACACCTTAGTGATGCGG 22519  
 226 ..... 226  
 22520 GACACTAAGAAATAGGCACTTTAGACAAATCGACTGGCAGGGATGGGA 22569  
 226 ..... 226  
 22570 GCCGGGAGAGAGGCTGTGGATGAGAGAGAAGAAATCATTTGGGAGCTA 22619  
 226 ..... 226  
 22620 CACACATCTGTGTGGCGCTGCACACTGAGTCTGTGGAACTCAAGTGGGCT 22669  
 227 ..... GlyAspSerGlyGlyProLeuVal 234  
 22670 CTGAGGCTGACTGCTGCCCTCCCGCAGGGGAGCTCGGGCGGCCCTGGGT 22719  
 234 lCysLeuValGlyGlnSerTrpLeuGlnAlaGlyValIleSerTrpGlyG 251  
 22720 GTGCTCTGTGGGTCACTGTGTGCGCAGCGGGGGGTGATCAGCTGGGGTG 22769  
 251 lGlyGlyAlaArgGlnAsnArgProGlyValTyrIleArgValThrAla 267  
 |||||



55 snGlySerHisPheCySGlyGlySerLeuLeuAlaIuGlnTrpValLeu 71  
 |||||||  
 5038 ACGGAGCCACTTCTGCGGGGAGCCCTCATCGCGAGACAGTGGGCTCG 5087  
 72 ThrAlaHisCyPhe..... 77  
 |||||||  
 5088 ACGGCGCGCACTCTTCCCAAGTAGTCCGCCGCCCTGCCCCCGCC 5137  
 77 ..... 77  
 5138 CATAGCGTACAGCGCCCCCGCGCTACCGGTTACAGCACCGTGACAGC 5187  
 77 ..... 77  
 5188 GCCCGCGGCCAAATCTGCGGGTAGCTTCTGTGGGGCTCTGTGTCA 5237  
 77 ..... 77  
 5238 GCCCTCCCAACAGATGCTTCCCTTAGGTCCAACTCCAGGGCTAACTTC 5287  
 77 ..... 77  
 5288 CAGTTGCACACCGCTGTCTCCGCGCGGAGAGTGCCTGCACCGTCCCC 5337  
 77 ..... 77  
 5338 GACCCCTTCATCTCTCCACCCACTACCACTCTCTGTGGTCCCTGC 5387  
 77 ..... 77  
 5388 AAGACGGCGCGCGGCTCTGCCACCGGCCCTGTGGCTTCCCA 5437  
 77 ..... 77  
 5438 TCCCGCACACACTCAGTCCAGACACTCTTCCGGAGAGAACTCTGCT 5487  
 77 ..... 77  
 5488 TAGAAGCCCAAGACAGACAGAAAGGCCCTTCTCTCCCTCACCACT 5537  
 77 ..... 77  
 5538 GAACACCCCAAGAAAGCCCTGAGAGAGCCAGGCCACCCAGCCCTGC 5587  
 77 ..... 77  
 5588 CATGTATGACCACTGTGCGCCACACTTTCGGGTGTCCAGGCCCTCA 5637  
 77 ..... 77  
 5638 CCTCACACTCAGCAGCGCAGCTTAATTATTAAACCCACATCTTTT 5687  
 77 ..... 77  
 5688 TCTTTTTTCTCTTGATCTTTAAAGATATCATGACAAAAAACCC 5737  
 77 ..... 77  
 5738 CACATCTTAATTCAGATATCAGGCGCAGCGGTGCTCACACCGCT 5787  
 77 ..... 77  
 5788 AATCCACACTTTGGAGGCCAAGCGGCGAGATCAGTTGAGCCAGGA 5837  
 77 ..... 77  
 5838 GTTCAAGACAGCCCGGCAACAGACAGACCTGTCTACTAAAT 5887  
 77 ..... 77  
 5888 AAAAAAAAAATTAACAAAAACAAAAAACCCAGGTGAGTG 5937  
 77 ..... 77

5938 GTGCTGTGTTCCCAAGTACTCAGGAAGCTGAGCCAGAGAGGCTTGA 5987  
 77 ..... 77  
 5988 GCCAGAGATTGAGAGCTGAGTAGGCTACAGTTGCACACTGCATCCA 6037  
 77 ..... 77  
 6038 GCTGGGTGACAGAAATGAATCTGTCTTAGGGGAAAAAATGGA 6087  
 77 ..... 77  
 6088 AACTGGGCTAGAGAGAGGAGGTACCTGCCAGAGCCCGCAGGTGGG 6137  
 77 ..... 77  
 6138 CTGGCCCTGTGCACCTGAGAGGTGTCTCTCCGAGCAGCCTTTCAGG 6187  
 77 ..... 77  
 6188 TAGCACTGTACTGTCCCAACAGACAGAGTGAGAAAGTGCATACAGA 6237  
 77 ..... 77  
 6238 GAAGTTGTTAGTGACTTGCACAGGCCACACAGCGGCCAGGGGTGAGG 6287  
 77 ..... 77  
 6288 TGAATTCAGAGTGGCGGCTGTGCTTAATGCAACACTCATACCCG 6337  
 77 ..... 77  
 6338 TGCACCTGGCTGTCTTAGATGCTCTGAAAGACTGTGACAGAGAG 6387  
 77 ..... 77  
 6388 GTTGGGGCAAGTGTGGCAGAGAGTCCCGCCCTCCGGGCTGT 6437  
 78 ..... Arg. AsnT 80  
 6438 CCCCAACAGAGGGCCGCTACAGCCGCTCTGTCTTCTCCGCGACGA 6487  
 80 hrSerGluThrSerLeuTyrGlnValLeuGluAlaArgVal 96  
 |||||||  
 6488 CTTGTGAGAGTCTCTTACAGGCTCTGCTGGGGCAAGGCACTAGTG 6537  
 97 GluProGlyProHisAlaMetTyrAlaArgValArgGlnValGluSer 113  
 |||||||  
 6538 CAGCGGGACACACGCTATGTATGCCGGGTGAGGGGTGAGAGCA 6587  
 113 nProLeuTyrGlnGlyThrAlaSerSerAlaAspValAlaLeuValGlu 130  
 |||||||  
 6588 CCCCTGTACAGGGCATGGCTTCCAGCCGTGACGTGGCTGTGAGAG 6637  
 130 euGluAlaProValProPheThrAsnTyrIleLeuProValCysLeuPro 146  
 |||||||  
 6638 TGGAGGACCAAGTCCCTTACCAATTACATCTCCCGGTGCTGCT 6687  
 147 AspProSerValIlePheGluThrGlyMetAsnCysTyrValThrGly 163  
 |||||||  
 6688 GACCCCTCGGTGATCTTTGAGAGCGGCATGACTGTGGGTACTGCTG 6737  
 163 pGlySerProSerGlu..Glu..... 169  
 |||||||  
 6738 GGGAGCCCAAGTAGAGAGTAAAGGACAGGGCTGGGAAGATGGGG 6787  
 169 ..... 169  
 6788 GATGTGCTCAAGAGAGCCAGAGCGCTTGGGACCACTGTGAACCCAG 6837  
 169 ..... 169

```

6838 AATCTCTGCTGCTATTAATACATACACGGCTCTGTTACACAG 6887
169 ..... 169
6888 GTGGGGCCGAGGGAGCAGCAGACCCAGAAAGAGGGGTGAAG 6937
169 ..... 169
6938 GAGAGGGAGAGAAAGGGCCAGCTGCTGGAAGAGGAGATGT 6987
169 ..... 169
6988 CTGCCCCACAGTGCCTTAAGCCAGATGCTAAACATGAGAGAGAC 7037
169 ..... 169
7038 AGCAGAGTGGAGTGTGTCCCGGGGCGGTCTTGATGAGAGAG 7087
170 ..... AspleuLeuProgluP 175
7088 TCCGTTAGCCCTGGGCTGTTCGCCCTCCAGACCTCTCCGCGAAC 7137
175 roArgileuGlnLysLeuAlaValProIleAspThrProLysCys 191
|||||
7138 CGCGATCTCCAGAACTGCTGTGCCATTCAGACACCAAGTGC 7187
192 AsnLeuLeuTyrSerLysAspThrGluPheGlyTyrGlnProLysThrI 208
|||||
7188 AACCTGCTACACCAAGACACCGAGTTGGCTACCAACCCAAACAT 7237
208 elyAsnAspMetLeuCysAlaGlyPheGluGlyLysLysAspAlaC 225
|||||
7238 CAAGATGACATGCTGTGCGCGCTTCGAGAGGCAAGAGATGCT 7287
225 ySLys..... 226
7288 GCAAGGTGGAGCGATGTGTCCATGGGAGTCACTCTCACCGCAG 7337
226 ..... 226
7338 GCCCAGGGCAGGTGGATCATGCCCGCTCATGCGTATGCTTCAAC 7387
226 ..... 226
7388 CAGTGAACACCTTGTGTTGAAGAGCAGATTCCAGGGCGGCGGTG 7437
226 ..... 226
7438 GCTACGCCCTGTAAATCCAGCACTTGGGAGGCAAGAGGGTGGATCAC 7487
226 ..... 226
7488 AAGGTGAGAGATGAGACCATCTGGCTAACACAGTGAACCCCGTCTC 7537
226 ..... 226
7538 TACTTAAATAACAAAAAATTAAGCCAGCGGTGTGGCGCTGTATGTC 7587
226 ..... 226
7588 CCAGCTACTCGGAGGCTGAGGAGCAGAGAGATGGCTGAACCTGGAGGCG 7637
226 ..... 226
7638 GAGCTTGCAGTGAAGCAGATCGCCGCACTGCTCAAGCTGGGCCACA 7687
226 ..... 226
7688 GAGCAAGATTCGCTCAAAAAAAAAAAAAAAAAAGAAAGACAGATTCC 7737
226 ..... 226
7738 TGGACCCACCCAGAGAGATCAGATCAGTGGGTCTGGGGTGGGCACA 7787

```

```

226 ..... 226
7788 AGAACTGCATTTTCACACACACACACACCTAGTTGATCTGGGAGAC 7837
226 ..... 226
7838 TAAAGATATGACCTTTAGACAATCGACTGGCAGGGAGCGGAGCCG 7887
226 ..... 226
7888 GCAGAGAGCTCTGGAGATGAGAGAGAAATTCATTGTGAGCTACACGC 7937
226 ..... 226
7938 CTCTGTGCGCTGCACTGAGTCTCTGGAACTCAAGTGGGCTCTGAC 7987
227 ..... GlyAspSerGlyLysProLeuValLysL 236
7988 GGTGATTTGCGTCCCGCCAGGCGGAGCTCGGGCGGCGCTGTGTGCC 8037
236 euValGlyInserThrLeuGlnAlaGlyValIleSerTrpGlyGlu 252
|||||
8038 TCGTGTGCTAGTGTGCTGCGAGCGCGGTGATCAGCTGGGTGAGGCG 8087
253 CysAlaArgGlnAsnArgProGlyValTyrIleArgValThrAlaHisH 269
|||||
8088 TGTGCGCGCCAGACCGCCAGGTGTTATCATCGCTGACCGCCACCA 8137
269 sAsnTrpIleHisArgIleIleProLysLeuGlnPheGlnProAlaArgL 286
|||||
8138 CACTGTGATCATCGATCATCCCAAACTGCAAGTCCAGCCAGCAAGT 8187
286 euGlyGlyGlnLys 290
|||||
8188 CGGGCAGCCAGAG 8201

seq_name: gb_pat:AX001347
seq_documentation_block: 683 bp DNA linear PAT 10-MAR-2000
LOCUS AX001347
DEFINITION Sequence 3 from Patent EP0890646.
ACCESSION AX001347
VERSION AX001347.1 GI:7241523
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 683)
AUTHORS
Burgess,N.A. and Southan,C.D.
TITLE
HE2NM40 Serine Protease
JOURNAL
Patent: EP 0890646-A 3 13-JAN-1999;
SMITHKLINE BEECHAM PLC (GB)
FEATURES
source
1..683
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 156 a 205 c 195 g 119 t 8 others
ORIGIN
alignment_scores:
quality: 810.50 length: 220
ratio: 4.243 gaps: 7
Percent Similarity: 86.818 Percent Identity: 79.091
alignment_block:
US-10-041-006-7 x AX001347
Align seg 1/1 to: AX001347 from: 1 to: 683
78 ArgAsnThrSerGluThrSerLeu.TyrGlnValLeuLeuGlyAlaArg 94
|||||

```

```

1 CGCAACACCTGTGAGACGTCCCTGGACACAGTCTCTGGGGGCAAGGC 50
94 InLeuValGlnPro.GlyProHisAlaMetTyrAlaArgValArgGlnVal 110
51 AGCTAGTGCAGCCGGGGACACACACCTGTATGTATGCCGGGTGAGGCAAGT 100
110 IGluSerAsnProLeuTyrGlnGlyThrAlaSerSerAlaAspValAlaL 127
101 GGAGAGCAACCCCTGTACAGAGGCGCCCTCCAGCCGTGACGTGGCC 150
127 euValGluLeuGlnAlaProValProPheThrAsnTyrIleLeuProVal 143
151 TGGTGAGCTGGAGGACACAGTCCCTTACCAATTACATCTCTCCCGTG 200
144 CysLeuProAspProSerValIlePheGluThrGlyMetAsnCysTyrVa 160
201 TGCCTGCTGACCCCTCGGTGATCTTTGAGACGGGCATGAACCTGTGGGT 250
160 LThrGlyTyrGlySerProSerGluGluAspLeuProGluProArgI 177
251 CACTGCTGGGGGACGCCAGTGAAGAACCTCTGCGCCGACCGCGGA 300
177 ILeuGlnIlyLeuAlaValProIleIleAspThrProLysCysAsnLeu 193
301 TCCCTCAGAACTCCGTGCTGCCATCATCGACACCCAGTGCACACTG 350
194 LeuTyrSerLysAspThrGluPheGlyTyrGlnProLysThrIleLysAs 210
351 CTCTACAGCAAGACACCGAGTTTGGCTACCAACCAACCATTAAGAA 400
210 nAspMetLeuCysAlaGly.PheGluGlnGlyLysLysAspAlaCysLys 226
401 TGACATGCTGTGCGCGGNTTTCAGAGAGGCAAGAGATGCCCTGCAG 450
227 .GlyAspSerGlyGlyProLeuValCysLeuValGlyGlnSerTyrLeu 243
451 TGGG.....GGAGCCATGTGGTGTCCATGGGATTCAGTTCACAC 491
243 InAlaGlyVal...IleSerTyrGlyGlnGly.....Cys 253
492 GCNCAGGCCGACGAGGAGGTGGGTAAAGCCCTGTCCATGGGTATG 541
254 AlaArgIlyAsnArgProGlyValTyrIleArgValIThrAlaHisLys 270
542 CTTCAAACAGAACGCCGAGTGTAAATCCGTGTAAGGCCACAA 591
270 nTrrPheIleAsnArgIleIleProLysLeuGlnPheGlnProAlaArgLeu 287
592 CTGGATCCATCGATCATATCCCAACTGAG..TTCAGGCCAACGAAGTTTG 640
287 Iy 287
641 GC 642

```

seq\_name: gb\_pat:E31783

seq\_documentation\_block:

LOCUS E31783

683 bp

DNA

linear

PAT 07-FEB-2001

DEFINITION Novel compound.

ACCESSION E31783.1 GI:13026587

VERSION JP 199098992-A/2.

KEYWORDS unidentified.

SOURCE unidentified.

ORGANISM unclassified.

REFERENCE 1 (bases 1 to 683)

AUTHORS Nicols,A.B.H.H. and Kurinkembialdo,C.D.S.S.

TITLE Novel compound

JOURNAL Patent: JP 199098992-A 2 13-APR-1999;

COMMENT SMTHTLINE BECHAM CORP PUBLIC LTD CO

OS Unidentified

PN JP 199098992-A/2

PD 13-APR-1999

```

PF 09-JUN-1998 JP 1998199425
PR 10-JUN-1997 GB 9712088.5,17-OCT-1997 GB 97308295-1, PR
20-FEB-1998 GB 9803650.2
PI NICOLA ANN BAJESU HELEN ELIZABETH KURINKENBIADO, PI
CHRISTOPHER DONALD SAZAN
PC C12N15/09,A61K31/70,A61K35/76,A61K38/00,A61K38/00,
PC A61K39/395,
PC A61K39/395,A61K48/00,C07K16/40,C12N5/10,C12P21/02,C12P21/08,
PC C1201/68,
PC C12N15/00,A61K37/02,A61K37/02,A61K37/02,C12N5/00 CC
Strandedness: Single;
CC Topology: linear;
FH Key Location/Qualifiers
FT source 1..683
Location/Qualifiers
source 1..683
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 156 a 205 c 195 g 119 t 8 others
ORIGIN

```

```

alignment_scores:
  Quality: 810.50 Length: 220
  Ratio: 4.243 Gaps: 7
Percent Similarity: 86.818 Percent Identity: 79.091

```

alignment\_block:

US-10-041-006-7 x E31783

Align seg 1/1 to: E31783 from: 1 to: 683

```

78 ArgAsnThrSerGluThrSerLeu.TyrGlnValIleuLeuGlyAlaArg 94
|||||
1 CGCAACACCTGTGAGACGTCCCTGGACACAGTCTCTGGGGGCAAGGC 50
94 InLeuValGlnPro.GlyProHisAlaMetTyrAlaArgValArgGlnVal 110
51 AGCTAGTGCAGCCGGGGACACACACCTGTATGTATGCCGGGTGAGGCAAGT 100
110 IGluSerAsnProLeuTyrGlnGlyThrAlaSerSerAlaAspValAlaL 127
101 GGAGAGCAACCCCTGTATACAGAGGCGCCCTCCAGCGCTGAGGTGGCC 150
127 euValGluLeuGlnAlaProValProPheThrAsnTyrIleLeuProVal 143
151 TGGTGAGCTGGAGGACACAGTCCCTTACCAATTACATCTCTCCCGTG 200
144 CysLeuProAspProSerValIlePheGluThrGlyMetAsnCysTyrVa 160
201 TGCCTGCTGACCCCTCGGTGATCTTTGAGACGGGCATGAACCTGTGGGT 250
160 LThrGlyTyrGlySerProSerGluGluAspLeuProGluProArgI 177
251 CACTGCTGGGGGACGCCAGTGAAGAACCTCTGCGCCGACCGCGGA 300
177 ILeuGlnIlyLeuAlaValProIleIleAspThrProLysCysAsnLeu 193
301 TCCCTCAGAACTCCGTGCTGCCATCATCGACACCCAGTGCACACTG 350
194 LeuTyrSerLysAspThrGluPheGlyTyrGlnProLysThrIleLysAs 210
351 CTCTACAGCAAGACACCGAGTTTGGCTACCAACCAACCATTAAGAA 400
210 nAspMetLeuCysAlaGly.PheGluGlnGlyLysLysAspAlaCysLys 226
401 TGACATGCTGTGCGCGGNTTTCAGAGAGGCAAGAGATGCCCTGCAG 450
227 .GlyAspSerGlyGlyProLeuValCysLeuValGlyGlnSerTyrLeu 243
451 TGGG.....GGAGCCATGTGGTGTCCATGGGATTCAGTTCACAC 491

```

```

243 InAlaGlyVal...IleSerTrpGlyGly...Cys 253
      ::::|
492 CGNCAGCCAGGCGAGGCGGTGTAAGGCCCTGTCATGGGTGTG 541
      ::::|
254 AlaArgGlnAsnArgProGlyValTyrIleArgValThrAlaHisHisAs 270
      ::::|
542 CTTCAAGAGGAGAACGCCAGGTCTTAAATCCGTGTAAGCGCCCAACAAA 591
      ::::|
270 nTrpIleHisArgIleIleProLeuLeuGlnPheGlnProAlaArgLeu 287
      ::::|
592 CTGGATTCATCGATCATNCCCAACTGAG.TTCCAGCCACGAAAGTTTG 640
      ::::|
287 Ty 287
      ::::|
641 GC 642

seq_name: gb_pr:AB010779
seq_documentation_block:
  title: 1306 bp mRNA linear PRI 17-MAY-2001
  accession: AB010779
  version: AB010779.1 GI:12248750
  keywords: Homo sapiens brain cDNA to mRNA.
  source: Homo sapiens
  organism: Homo sapiens
  reference: Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  title: Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
  authors: Mitsuhashi, S., Okui, A., Komiyama, K. and Yamaguchi, N.
  journal: Cloning and characterization of a human brain-specific serine
  title: protease, hbssp-4
  reference: 2 (bases 1 to 1306)
  authors: Yamaguchi, N. and Mitsuhashi, S.
  journal: Direct Submission
  title: Submitted (27-JAN-1998) Nozomi Yamaguchi, Institute for Geriatrics,
  journal: Kyoto Prefectural University of Medicine, Department of Cell
  journal: Biology, Kajicho 465, Kawaramachi Hirokoji, Kyoto,
  journal: Kyoto 602-8566, Japan (E-mail: nozomi@koto.kpu-m.ac.jp,
  journal: Tel:81-75-251-5848(ex.5848), Fax:81-75-251-5848)
  features:
    source:
      1..1306
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /tissue_type="brain"
      4..957
      /gene="hbssp-4"
      4..957
      /gene="hbssp-4"
      /note="prosemin"
      /codon_start=1
      /product="brain-specific serine protease-4"
      /protein_id="BAB20263.1"
      /db_xref="GI:12248751"
      /translation="MVFSGAPPALGGGCLGFTSTLLASTAIIINAARIPYPACGKP
      OOLNRYVGEDSTDESEWPWIVSIQKNGTHHCAGSLTSTVTAACEPDNKNRYLF
      SYLGAOMLGNGSRGSKQYVAMVEPHYVSMKEGACADIAIVLERSIOSERLPI
      CLPDAIHLPPNTHCWLISGWSIIOGCVPLPHQOTLOKLVPIIDEVSGSHLYWRAGQ
      GPTTDMLCAGTIEGERDCLDSDGGLPCQYDGAWLGLGIIISWEGCAERNRPVYI
      SLASHSWEKIVQGVYLRGAOGGALRALRAPSGSGAARS"
    gene:
      cds:
        alignment_scores:
          quality: 739.00 length: 285
          ratio: 3.421 gaps: 4
          percent similarity: 75.789 percent identity: 48.772
        alignment_block:
          US-10-041-006-7 x AB010779

```

```

Align seg 1/1 to: AB010779 from: 1 to: 1306
10 LeuLeuLeuLeuCys.....PheGlySerGlnArgAlaLysAl 22
      ::::|
64 CTGCTGCTGCTGGCGACGACGACCATCTCAATGCGCGCATTCACCTGTT 113
      ::::|
22 aaAlaThrAlaCysGlyArgProArgMetLeuAsnArgMetValGlyGly 39
      ::::|
114 TCCGCCAGCCTGTGGAGAGCCCGACGACGTAACCGGTTGTGGCGCGG 163
      ::::|
39 InAspThrGlnGlyGlyGlyTyrProTyrProTyrGlnValSerIleGlnArgAsn 55
      ::::|
164 AGGACAGCACTGACAGCAGACTGCCCTGATGCTGATGATCCAGAGAAAT 213
      ::::|
56 GlySerHisPheCysGlyGlySerLeuIleAlaGlnGlnTyrValLeuH 72
      ::::|
214 GGAACCCACACCTGCGCAGAGTTCTGCTCACCAGCGCGTGGGTGATCAG 263
      ::::|
72 AlaAlaHisCysPheArg...AsnThrSerGluThrSerLeuTyrGln 88
      ::::|
264 TGCTGCCCATCTGTTCAAGACACCACTGAACAACCATACCTGTTCTGTC 313
      ::::|
88 alLeuLeuGlyAlaArgGlnLeuValGlnProGlyProHisAlaMetTyr 104
      ::::|
314 TGCTGCTGGGGCGCTGGAGACCTGGGAGACCTGCTGCTGCTGCCAGAG 363
      ::::|
105 AlaArgValArgGlnValGlnSerAsnProLeuTyrGln...GlyThrAl 120
      ::::|
364 GTGCGTGTGCTGCTGGGTGAGCCCGACCCCTGCTATTCCTGGAAGACAG 413
      ::::|
120 aserSerAlaAspValAlaLeuValGlnLeuGlnAlaProValProPheT 137
      ::::|
414 TGCTGTGCGACATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 463
      ::::|
137 hrasnTyrIleLeuProValCysLeuProAspProSerValIlePheGln 153
      ::::|
464 CAGAGCGGGGCTGCGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 513
      ::::|
154 ThrGlnMetAsnCysThrValThrGlyTyrGlySerProSerGlnGln 170
      ::::|
514 CCAACACCCACCTGCTGATCTGAGCTGGGGAGGATCCAAAGATGAGCT 563
      ::::|
170 pLeuLeuProGluProArgIleLeuGlnLysLeuAlaValProIleLea 187
      ::::|
564 TCCTTGCCGCCACCTGACAGCCCTGACAGAGCTGAAGTTCATCATGATG 613
      ::::|
187 spThrProLysCysAsnLeuLeuTyrSerLysAspThrGlnPheGlyTyr 203
      ::::|
614 ACTCGGAAGTCTGACACCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
      ::::|
204 GlnProLysThrIleLysAsnAspMetLeuCysAlaGlyPheGlnGln 220
      ::::|
661 .....CCCATTCACCTGAGGACATCTGCTGCTGCTGCTGCTGCTGCTG 701
      ::::|
220 yLysLysAspAlaCysLysGlyAspSerGlyGlyProLeuValCysLeuV 237
      ::::|
702 GGAGCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 751
      ::::|
237 alGlyLeuSerThrLeuGlnAlaGlyValIleSerThrPheGlyGlyLys 253
      ::::|
752 TGGACGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 801
      ::::|
254 AlaArgGlnAsnArgProGlyValTyrIleArgValThrAlaHisHisAs 270
      ::::|
802 GCGGAGCGGACAGGCGCGGCTTACATCAGCTCTCTGCGCGCGCGCTC 851
      ::::|
270 nTrpIleHisArgIleIleProLeuLeuGlnPheGlnProAlaArgLeu 287
      ::::|
852 CTGGTGAAGAGTGTGTCAGAGGGGTGACGCTCGCGGGCGCGCTCAGG 901
      ::::|
287 TyGly 288
      ::::|

```







Wed Aug 14 08:39:13 2002

us-10-041-006-7.p2n.rge

Page 21









[illegible]







(HYSE-) HYSEQ INC.

P-PSDB; AAM41174.

Claim 1; SEQ ID NO 4319; 10078bp; English.

specification.

Sequence 1157

Quality: 1522.00

Similarity: 98.630

US-10-041-006-7 x AAI60330

Align seg 1/1 to: AA16033

190 GAATGGTGGGCGGCCAGG

KW Serine p:

KW BCOM3; testisin; fertility; suppressor; testicular germ cell cancer  
KW seminoma; testis-specific expression; antitumour; sperm development;  
KW infertility; human; chromosome 16p13.3; ss.

OS Homo sapiens.

Key	Location/Qualifiers
EH	
FT	
CDS	3. .899

```
FT      /product= SP003LA
FT      /note= "sequence contains one internal stop codon"
..
```

(AMRA - ) AMRAD OPERATIONS PTY LTD.

Antalis TM, Hooper JD,

DR WPI; 1998-480768/41.  
DR P-PSDB; AAW77304.

PR New serine protease(s) and kinase involved in regulating cell  
 PR activity and viability - particularly the testis-specific protease  
 PR Hs2A2 used for modulation of fertility and as tumour suppressor  
 XX  
 PS Example 15; Fig 20C; 167pp; English.

CC AAV59134-36 represent HELA2 homologues. The genes are found in a cluster  
CC on chromosome 10p13.3. HELA2 was isolated from HeLa cells, and has  
CC homology to serine proteases. The protein is involved in or associated  
CC with regulation of cell activity and/or viability. Administration of  
CC recombinant HELA2 (also called testisin) is used to increase fertility.  
CC Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of  
CC testicular germ cell cancers (seminoma) and is also expressed in some  
CC non-testicular cancers (of colon, pancreas, prostate and ovary), so is  
CC a marker/potential therapeutic target for cancer. The promoter from the  
CC HELA2 gene is useful for testis-specific expression of other genes,  
CC e.g. for gene therapy or modulation of fertility. Drugs that block  
CC activity of HELA2 should have antitumour activity (other than in  
CC testis) while in testis recombinant HELA2 should stop growth of tumours  
CC and normalise sperm development (eliminating the need for orchidectomy).  
CC Identification of mutant forms of HELA2 can be used to diagnose  
CC infertility.

Sequence 980 BP; 192 A; 325 C; 289 G; 174 T; 0 other;

```

alignment_scores:
  Quality: 1449.00      Length: 265
  Ratio: 5.468          Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

alignment\_block:

Align seg 1/1 to: AAV59136 from: 1 to: 980

26 CysGlyIvGPProAlaMetLeuAsnArgMetValIlglyIvGlnAspTrpIgl 42  
3 TGTGGTCCGCCAGGATCTTAACCGAATGGTGGCGGCGACGACAGCA 52  
42 ngluIglvIvtrpProtrpGlnValSerIleGlnArgAsnIlySerHisP 59  
53 GGAGGGCCAGATGGCCCTGGCAAGTACGATCCACGCGCAACGGAAGCCACT 102  
59 heCysGlyIvSerIleuIlealacIvIvtrpValIleuThrAlaAlaHis 75

[illegible]







DR PFI: 1999-072882/07.  
 DR P-PSDB; AAM88454.  
 PT New serine protease HE2NM40 polypeptides and polynucleotides -  
 PT useful as diagnostic reagents and for prevention and treatment of  
 PT cancer, Alzheimer's disease and asthma  
 PS Claim 18; Page 18-19; 21pp; English.  
 XX  
 CC This is the nucleotide sequence of an expressed sequence tag (EST)  
 CC clone. A polypeptide encoded by the clone is provided in AAM88454.  
 CC Claimed polynucleotides may comprise a nucleotide sequence which  
 CC has at least 70% identity to this sequence. Such polynucleotides  
 CC can be obtained from a cDNA library derived from mRNA in cells of  
 CC human neutrophils, adipose or synovial tissue, using expressed  
 CC sequence tag analysis. A polynucleotide encoding a serine protease  
 CC termed HE2NM40 (see W884453) is specifically claimed. HE2NM40  
 CC polypeptides and polynucleotides are useful for diagnosing diseases  
 CC related to over or underexpression of HE2NM40 protein. They can  
 CC also be used to modulate HE2NM40 activity. Diseases diagnosed,  
 CC prevented or treated include cancer, inflammation, asthma, wasting  
 CC diseases, atherosclerosis, stroke, diabetes, arthritis, and  
 CC neurodegenerative diseases and disorders including Alzheimer's  
 CC disease.  
 XX  
 SQ Sequence 683 BP; 156 A; 205 C; 195 G; 119 T; 8 other;  
 XX  
 alignment\_scores:  
 Quality: 810.50 Length: 220  
 Ratio: 4.243 Gaps: 7  
 Percent Similarity: 86.818 Percent Identity: 79.091  
 alignment\_block:  
 US-10-041-006-7 x AAX06943 ..  
 Align seg 1/1 to: AAX06943 from: 1 to: 683  
 78 ArgAsnThrSerGluThrSerLeu.TyrGluValIleuLeuGlyAlaArg 94  
 1 CGCAACACCTCTGAGACCTCCCTGACACAGCTCTGCTGGGGCGAAGC 50  
 94 InLeuValGlnPro.GlyProHisAlaMetTyrAlaArgValArgGlnVa 110  
 51 AGCTAGTCACAGCGGGGACACACGCTATGATGCGCGGAGAGCAGGT 100  
 110 IGluSerAsnProLeuTyrGlnGlyThrAlaSerSerAlaAspValAlaL 127  
 101 GGAGAGCAACCCCTGTGTACAGGCGACGCTCCAGCGCTGACGTGGGCC 150  
 127 euValGluLeuGluAlaProValProPheThrAsnTyrIleLeuProVal 143  
 151 TGTGTGAGACTGAGGACCACTGCTTACCAATACATCTCCCGCTG 200  
 144 CysLeuProAspProSerValIlePheGluThrGlyMetAsnCysTyrPva 160  
 201 TGCGTGCTGACCCCTGCTGATCTTTGAGACGGGCAATGAAGACTCTGGGT 250  
 160 1ThGlyTyrPglYserProSerGluGluAspLeuLeuProGluPrcArgI 177  
 251 CACTGGCGTGGGGGACCCCAATGAGGAGACCTCTGCCCAACCGGCGGA 300  
 177 leLeuGlyLeuAlaValProIleIleAspThrProLysCysAsnLeu 193  
 301 TCTGTGAGAAATCGCTGCTGCCCATCATCGACACACCAAGTGCACACTG 350  
 194 LeuTyrSerLysAspThrGluPheGlyTyrGlnProIlyThrIleLysAs 210  
 351 CTCTACAGCAAAAGACACCGAGTTGGTACCAACCAAAACATCAAGAA 400  
 210 nasPmetLeucylsalagly.PhegluGluGlyLysAspAlaCysLys 226  
 401 TGACATGCTGTGCGCGGNTTTCAGAGGAGGCGCAAGAGATGCTGCAAG 450

227 GLYAspSerGlyGlyProLeuValCysLeuValGlyGlnSerTrpLeuG 243  
 451 TGGG.....GGACCATGTGTGTTCCATGGGAGATTCAGTTCCAAAC 491  
 243 InAlaGlyVal...IleSerTrpGlyGluGly.....Cys 253  
 492 CGNACAGGCCAGGGCAAGGTGGGTTAAGCCCTGTCCATGGGTGATG 541  
 254 AlaArgGlnAsnArgProGlyValTyrIleArgValThrAlaHisHis 270  
 542 CTTCACAAACAAACAGCCAGTGTAAATCCGTGTAACGGCCACCAAA 591  
 270 nTrpIleHisArgIleIleProLysLeuGlnPheGlnProAlaArgLeu 287  
 592 CTGTGATTCATCGATCATNCCAACTGAG.TTCCAGCCCAAGAGTTTG 640  
 287 Ly 287  
 641 GC 642  
 seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA61696  
 seq\_documentation\_block:  
 ID AAA61696 standard; cDNA; 1007 BP.  
 AC AAA61696;  
 XX  
 DT 23-OCT-2000 (first entry)  
 XX  
 DE cDNA encoding human serine protease BSSP4 (hbssp4) SEQ ID NO:3.  
 XX  
 KW BSSP4; serine protease; human; hbssp4; mouse; mbssp4; brain;  
 KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;  
 KW oedema; dropsy; cancer; inflammation; prostate; testis; bone; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200031277-A1.  
 XX  
 PD 02-JUN-2000.  
 XX  
 PF 19-NOV-1999; 99WO-JP06472.  
 XX  
 PR 20-NOV-1998; 98JP-0347813.  
 XX  
 PA (FUSO ) FUSO PHARM IND LTD.  
 XX  
 PI Uemura H, Okui A, Kominami K, Yamaguchi N, Mitsui S;  
 XX  
 DR WPI: 2000-400084/34.  
 DR P-PSDB; AAB11701.  
 XX  
 PT Serine protease BSSP4 and antibodies recognizing BSSP4 for assay and  
 PT diagnosis of diseases in which BSSP4 expression is altered  
 XX  
 PS Claim 4; Page 67-69; 11pp; Japanese.  
 XX  
 CC The invention relates to novel serine proteases designated BSSP4  
 CC (AAB1700-B11709), and to nucleic acids encoding them (AAA61695-A61704,  
 CC AAA61799). The invention also relates to vectors and transformants  
 CC comprising BSSP4 nucleic acids; transgenic animals in which the  
 CC expression level of BSSP4 can be varied; and an mbssp4 knockout mouse.  
 CC The invention additionally encompasses anti-BSSP4 antibodies and methods  
 CC of production of such antibodies, methods of BSSP4 detection using the  
 CC antibodies, and the use of BSSP4 proteins or fragments as diagnostic  
 CC markers for certain medical conditions. Nucleotides encoding BSSP4 were  
 CC initially isolated in a human brain cDNA library using degenerate PCR  
 CC primers (AAA61714-A61715) based on conserved regions of serine  
 CC proteases. The BSSP4 serine proteases and nucleotides encoding them are  
 CC useful in detecting homologues, mutants and polymorphic variants in  
 CC biological samples (e.g., blood, urine, brain, prostate gland and testis)  
 CC as diagnostic markers for diseases associated with altered BSSP4

CC expression levels. Such diseases include Alzheimer's disease, oedema  
 CC (dropsy), cancer or inflammation of brain, prostate, testis or bone.  
 CC Sequences AAA61695-A61703 and AAA61799 represent cDNAs encoding human  
 CC BSSP4 variants (hBSSP4), and sequence AAA61704 represents cDNA encoding  
 CC murine BSSP4 (mBSSP4).

XX Sequence 1007 BP; 171 A; 311 C; 320 G; 205 T; 0 other;

#### alignment\_scores:

Quality: 739.00 Length: 285  
 Ratio: 3.421 Gaps: 4  
 Percent Similarity: 75.789 Percent Identity: 48.772

#### alignment\_block:

US-10-041-006-7 x AAA61696 ..

align seg 1/1 to: AAA61696 from: 1 to: 1007

```

10 LeuLeuLeuLeuGys.....PheGlySerGlnArgAlaGlyAla 22
   |||||||.....:|||||:|||||:|||||:|||||:|||||:
64 CTCTGCTGCTGGCGACGACGACGACCTCAATGCGCGCCAGGATACCTGT 113
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 TCCCCCAGCCTGTGGGAGGCCGACGACGACGACGCGGTGTGGGCGCG 163
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
39 LnaSPHrGlnGlnGlyGlnTPrProTPrGlnValSerIleGlnArgAsn 55
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 AGGACACACACTGACAGCAGAGTGGCCCTGGATCTGAGCATCCAGAGAA 213
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
56 GlySerIlePheGlySerIleGlnValGlnIleGlnIleProValLeu 72
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 GGGACCCACACTGCGAGGCTCTCTGCTACCAAGCCGCGGTGATGATC 263
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 rAlaAlaHisCysPheArg...AsnThrSerGlnThrSerLeuTyrGln 88
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 TGGTGGCCCACTGTTTCAAGACAACCTGACAAACCATACCTGTCTCTG 313
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
88 AlLeuLeuGlnAlaArgGlnLeuValGlnProGlyProHisAlaMetTyr 104
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 TGCCTGGGCGGCGCTGGGAGAACCTGCTGCTCGCTCCAGAACG 363
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
105 AlaArgValArgGlnValGlnSerAsnProLeuTyrGln...GlyThrAl 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 GTGGGCTTTCCTGGGTGGAGCCCACTGTGTATTCCTGGAGAGGAGG 413
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 aSerSerAlaSPyAlaLeuValGlnLeuGlnAlaProValProPheT 137
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 TGCCTGGGAGACATGCTCCCTGGGCTGTCGAGCGCTCATACAGTTCT 463
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
137 hIraSPHrIleLeuProValCysLeuProAspProSerValIlePheGln 153
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 CAGAGCGGCTCTGCCATCTGCTACCTGATGCTTATCCACCTCCCT 513
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 ThrGlyMetAsnCysTrpValThrGlyTrpGlySerProSerGlnGln 170
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
514 CCAACACACCACTCTGATCTGAGCTGGGGGAGCATCAAGATGAGT 563
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
170 PLeuLeuProGlnProArgIleLeuGlnValLeuAlaValProIleLea 187
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
564 TCCCTGGCCCACTGACAGACCTGACAGAGCTGAAGGTTCTTATCATCG 613
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 SPHrProLysCysAsnLeuLeuTyrSerLysAspThrGlnPheGlyTyr 203
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
614 ACTGGAAGCTGCAAGCATGTGACTGGCGGAGAGCAAGACAGGGA... 660
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 GlnProLysThrIleLysAsnAspMetLeuCysAlaGlyPheGlnGln 220
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
661 .....CCATCACTGAGAGACATGCTGTGTCGCGGCTACTTGGAGGG 701
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 YLysLysAspAlaLacLysLysGlyAspSerGlyGlyProLeuValCysLeu 237
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

702 GGAGCGGATGCTTGTCTGGCGGACCTCCGGGCCCCCTCATGTGCCAGG 751
237 aGlyGlnSerTrpLeuGlnAlaGlyValIleSerTrpGlnGlyCys 253
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
752 TGGACGGCGCTGCTGCTGGCGGACATCATCACTGTGCGGAGGCTGT 801
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 AlaArgGlnAsnArgProGlyValTyrIleArgValThrAlaHisHis 270
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
802 GCCGAGCCCAACAGGCCCGGGGTCTACATCAAGCTCTGTGCGACCGCTC 851
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
270 nTrpIleHisArgIleIleProLysLeuGlnPheGlnProAlaArgLeu 287
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
852 CTGGGTGAGAAGATCTGCAAGGGGTGTCAGCTCCGCGCGCGCTCAGG 901
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
287 YGly 288
   |||||||
902 GGGGT 906

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA541085
seq_documentation_block:
ID AA541085 standard; cDNA; 1352 BP.
XX
AC AA541085;
XX
DT 17-DEC-2001 (first entry)
XX
DE cDNA encoding novel human enzyme polypeptide #301.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotropic; anticoagulant; SS.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218299.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225265.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.

```

PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226272
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226688
PR	23-AUG-2000	2000US-0227182
PR	30-AUG-2000	2000US-0227009
PR	01-SEP-2000	2000US-0229824
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	05-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229359
PR	06-SEP-2000	2000US-0229531
PR	06-SEP-2000	2000US-0230437
PR	08-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231122
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0233197
PR	12-SEP-2000	2000US-0233368
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0233400
PR	14-SEP-2000	2000US-0233401
PR	14-SEP-2000	2000US-0233403
PR	14-SEP-2000	2000US-0233404
PR	14-SEP-2000	2000US-0233405
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236361
PR	29-SEP-2000	2000US-0236365
PR	29-SEP-2000	2000US-0236366
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239935
PR	13-OCT-2000	2000US-0239937
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241121
PR	01-NOV-2000	2000US-0241826
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246479
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610

PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249209
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249212
PR	17-NOV-2000	2000US-0249213
PR	17-NOV-2000	2000US-0249214
PR	17-NOV-2000	2000US-0249215
PR	17-NOV-2000	2000US-0249216
PR	17-NOV-2000	2000US-0249217
PR	17-NOV-2000	2000US-0249218
PR	17-NOV-2000	2000US-0249244
PR	17-NOV-2000	2000US-0249245
PR	17-NOV-2000	2000US-0249264
PR	17-NOV-2000	2000US-0249265
PR	17-NOV-2000	2000US-0249297
PR	17-NOV-2000	2000US-0249300
PR	17-NOV-2000	2000US-0249309
PR	01-DEC-2000	2000US-0250160
PR	01-DEC-2000	2000US-0250391
PR	05-DEC-2000	2000US-0251030
PR	05-DEC-2000	2000US-0251988
PR	05-DEC-2000	2000US-0256719
PR	06-DEC-2000	2000US-0251479
PR	08-DEC-2000	2000US-0251855
PR	08-DEC-2000	2000US-0251866
PR	08-DEC-2000	2000US-0251868
PR	08-DEC-2000	2000US-0251989
PR	11-DEC-2000	2000US-0251990
PR	11-DEC-2000	2000US-0254097
XX	05-JAN-2001	2001US-0259678
PA	(HUMA-)	HUMAN GENOME SCI INC.
PI	Rosen CA,	Barash SC, Ruben SM;
XX	WPI: 2001-465566/50.	
DR	P-PSDB: AAU23215.	
XX		
PT	Novel polypeptides and polynucleotides useful for diagnosing,	
PT	preventing, treating neural, immune system, muscular, reproductive,	
PT	pulmonary, cardiovascular, renal, proliferative disorders and cancerous	
PT	diseases	
XX		
PS	Claim 4: SEQ ID No 311; 1180pp; English.	
XX		
CC	The present invention relates to the isolation of novel human enzyme	
CC	polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences	
CC	encoding them. The enzyme polypeptides of the invention may comprise the	
CC	functional classes of oxidoreductases, transferases, hydrolases, lyases,	
CC	isomerases or ligases. The sequences of the invention are useful in the	
CC	diagnosis, treatment, prevention and/or prognosis of a wide range of	
CC	disorders including hyperproliferative disorders (e.g. cancer),	
CC	immunodeficiency disorders (e.g. AIDS) autoimmune disorders	
CC	(e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),	
CC	metabolic disorders (e.g. phenylketonuria), inflammatory disorders	
CC	(e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),	
CC	blood-related disorders (e.g. haemophilia), reproductive disorders	
CC	(e.g. infertility) and infectious disorders (e.g. Influenza). The	
CC	polynucleotides of the invention can also be used in gene therapy.	
CC	AAU20785-AAU21684 represent cDNA sequences encoding for the novel human	
CC	enzyme polypeptides of the invention.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at <a href="http://wipo.int/pub/published_pot_sequences">http://wipo.int/pub/published_pot_sequences</a> .	
XX		
XX	Sequence 1352 BP: 238 A: 446 C: 407 G: 261 T: 0 other:	





PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249266.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251858.  
PR 08-DEC-2000; 2000US-0251859.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI: 2001-476222/51.

P-PGDB: AAU17037.

Novel polypeptides and polynucleotides useful as diagnostic reagents to diagnose diseases or disorders associated with aberrant expression or activity of polypeptides, for treating blood clotting disorder, haemophilia

Claim 1; SEQ ID No 134; 601bp; English.

The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present

alignment\_scores:

Quality: 739.00 Length: 285  
Ratio: 3.421 Gaps: 4  
Percent Similarity: 75.789 Percent Identity: 48.772

alignment\_block:

US-10-041-006-7 x AAS26942









100

```

1 APPLICANT: BURESS, NICOLA A.
2 APPLICANT: CLINKENBEARD, HELEN E.
3 APPLICANT: SOUTHAN, CHRISTOPHER D.
4 TITLE OF INVENTION: NOVEL COMPOUNDS
5 NUMBER OF SEQUENCES: 6
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: RATNER & PRESTIA
8 STREET: P.O. BOX 980
9 CITY: VALLEY FORGE
10 STATE: PA
11 COUNTRY: USA
12 ZIP: 19482
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Diskette
15 COMPUTER: IBM Compatible
16 OPERATING SYSTEM: DOS
17 SOFTWARE: FastSeq for Windows Version 2.0
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/09/088,651
20 FILING DATE: JUNE 1, 1998
21 CLASSIFICATION:
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: GB9712088.5
24 FILING DATE: 10-JUNE-1997
25 APPLICATION NUMBER: EP 97308295.1
26 FILING DATE: 17-OCT-1997
27 APPLICATION NUMBER: GB 9803650.2
28 FILING DATE: 20-FEB-1998
29 ATTORNEY/AGENT INFORMATION:
30 NAME: PRESTIA, PAUL F.
31 REGISTRATION NUMBER: 23,031
32 REFERENCE/DOCKET NUMBER: GH30358
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE: 610-407-0700
35 TELEFAX: 610-407-0701
36 TELEX: 846169
37 INFORMATION FOR SEQ ID NO: 6:
38 SEQUENCE CHARACTERISTICS:
39 LENGTH: 1109 base pairs
40 TYPE: nucleic acid
41 STRANDEDNESS: single
42 TOPOLOGY: linear
43 MOLECULE TYPE: cDNA
44 US-09-088-651-6
45
46 alignment_scores:
47 Quality: 1169.50 Length: 295
48 Ratio: 4.515 Gaps: 5
49 Percent Similarity: 87.797 Percent Identity: 80.678
50
51 alignment_block:
52 US-10-041-006-7 x US-09-088-651-6 ..
53
54 Align seg 1/1 to: US-09-088-651-6 from: 1 to: 1109
55
56 1 MetAgtatgTgcAtaAlaAlaValProLeuenteuLeuLeuCysPheGlyse 17
57 198 ATGAGCGCGGGCCGGCGCGGTCCGCCCTCCTGCCTGCTGTTTNTNNGTN 24
58 17 tGlnatrgAlaLysAlaAlaIatrrAlaLacysGlyArgProArgrMetLeuAsnA 34
59 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
60 248 TCAGAAGGCCCAAGACGACGAACAACGTCGTGTGTCGCCGCCAGATGCTGAACN 29
61 34 rgmetValGlyGlyGlnAsprThrGlnGluGluGlyLuTrpProTrpGlnVal 50
62 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
63 298 NAANNtGcggNNngGcAgGAcACAGcAGcAGcAGcGAGTGcGcCCTGcCAAGTC 34
64 51 SerIleGlnArGuaGsnGlySerHisPheCysGlyGlySerIleAlaGl 67
65 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
66 348 AGCATCCAGCCCAAGcAGcAACcCANttTTTgCGGGGGcAGcTTTCAcGCGcGA 39
67 uGltTrpValIeuThrAlaAlaHisCysPheArGAsnThrSergLuThrS 84

```



```

|||||.....
398 GCAGTGGGTCTGAGGNTGCGACTGTTCCGCAACACTGTGAGAGT 447
84 eLeuTyGlnValLeuLeuGlnAlaArgGlnLeuValGlnProGlyPro 100
|||||.....
448 CCCTTACAGAGTCTGCTGGGGGCAAGGACGTAGTGCACCCGGGACCA 497
101 HisAlaMetTyAlaArgValArgGlnValGlnUserAsnProLeuTyGln 117
|||||.....
498 CACGNTATGATGCCCCGGGTGAGGAGGTGAGACCAACCCCTGTACCA 547
117 nglyThrAlaSerSerAlaAspValAlaLeuValGlnLeuGlnAlaProV 134
|||||.....
548 GGGCAGCGGCTTCAGCGCTGAGGTGGCCCTGGTGGAGCTGGAGCCAG 597
134 aLProPheThrAsnTyAlaLeuProValAlaCysLeuProAspProSerVal 150
|||||.....
598 TGGCCTTACCAATNACATCCCTCCCGTGGCTGCTGCTGACCCCTCGGTG 647
151 lIlePheGlnThrGlyMetAsnCysTrpValThrGlyTrpGlySerProse 167
|||||.....
648 ATCTTTGAGACGCGCATGAACTGTGGGTGCTGCTGGGGGAGCCGAG 697
167 rGlnGlnAspLeuLeuProGlnProArgIleLeuGlnIleuAlaValP 184
|||||.....
698 TAGGAGAACCTCTGCTGCCGAGACCCGGATCTGCAAGAACTGCTGTGC 747
184 rIleIleAspThrProLysCysAsnLeuLeuTySerLysAspThrGln 200
|||||.....
748 CCATCATGACACACCAAGTGCACACTGCTTACAGCAAAAGACACGAG 797
201 PheGlyTyGlnProLysThrIleLysAsnAspMetLeuCysAlaGlyP 217
|||||.....
798 TTTGGCTACCAACCCAAACCATCAAGAAATGACATGCTGTGGCGGNT 847
217 hseGlnGlnIlyLysAspAlaCysLys.GlyAspSerGlyGlyProle 233
|||||.....
848 TCGAGAGAGGCAAGAAAGATGCTGCAACTGGG.....GGAGCCAT 888
233 uValCysLeuValGlyGlnSerTrpLeuGlnAlaGlyVal...lIleSerT 249
|||||.....
889 GTGGTGTTCCTGAGGGGATTCAGTTCACACCGCAGCCGAGGCAAGGAT 938
249 rGlyGlnGly.....CysAlaArgGlnAsnArgProGly 260
|||||.....
939 GGGGTAAAGCCCTGTTCCATGGGTTATGCTTCAACAGAGAACGCCAGT 988
261 ValTyAlaArgValThrAlaHisAsnTrpIleHisArgIleIlePr 277
|||||.....
989 GTCTAATCGGTGTTNAAGGCCCAAAACGATGATCATCGATCATATNC 1038
277 oLysLeuGlnPheGlnProAlaArgLeuGly 287
|||||.....
1039 CAAACTGAG..TTCAGCCCAAGCAAGTTGGC 1068
seq_name: /sgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-088-651-3
seq_documentation_block:
; Sequence 3, Application US/0908651
; Patent No. 6165771
; GENERAL INFORMATION:
; APPLICANT: BURGESS, NICOLA A.
; APPLICANT: CLINKENBEARD, HELEN E.
; APPLICANT: SOUTHAN, CHRISTOPHER D.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P. O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/088.651
; FILING DATE: JUNE 1, 1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB9712088.5
; FILING DATE: 10-JUNE-1997
; APPLICATION NUMBER: EP 97308295.1
; FILING DATE: 17-OCT-1997
; APPLICATION NUMBER: GB 9803650.2
; FILING DATE: 20-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F.
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH30358
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 683 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-088-651-3

```

```

alignment_scores:
  Quality: 810.50      Length: 220
  Ratio: 4.243         Gaps: 7
  Percent Similarity: 86.818   Percent Identity: 79.091

```

```

alignment_block:
US-10-041-006-7 x US-09-088-651-3 ..

```

```

Align seg 1/1 to: US-09-088-651-3 from: 1 to: 683

```

```

78 ArgAsnThrSerGluThrSerLeu..TyGlnValLeuLeuGlnAlaArg 94
|||||.....
1 CGCAACACCTCTGAGACGTCCTCTGACACAGTCTGCTGGGGGCAAGC 50
94 InLeuValGlnPro.GlyProHisAlaMetTyAlaArgValArgGlnVal 110
|||||.....
51 AGCTAGTGCAGCGGGGACACAGCGTATGTATGCCGGGTGAGGAGGT 100
110 lGlnUserAsnProLeuTyGlnGlnTyThrAlaSerSerAlaAspValAla 127
|||||.....
101 GGAGAGCAACCCCTGTACAGGGGACGCGCTCCAGCGCTGACGTGGGCC 150
127 euValGlnLeuGlnAlaProValProPheThrAsnTyAlaLeuProVal 143
|||||.....
151 TGGTGGAGCTGAGGACACCAAGTGCCTTACCAATNACATCTCCCGTG 200
144 CysLeuProAspProSerValIlePheGlnThrGlyMetAsnCysTrpVal 160
|||||.....
201 TCCCTGCGCTGACCCCTCGGTATCTTGTAGACGGGCAAGAACTGCTGGGT 250
160 lThrGlyTrpGlySerProSerGlnGlnAspLeuLeuProGlnProArgI 177
|||||.....
251 CACTGCTGGGGGACCCCAAGAGAAAGACTCTGCGCCGACCGGCGGA 300
177 lLeuGlnIleuLeuAlaValProIleIleAspThrProLysCysAsnLeu 193
|||||.....
301 TCCCTGAGAAACTGCTGTCGCCATCATGACACACCAAGTGAAGCACTG 350
194 LeuTySerLysAspThrGlnPheGlyTyGlnProLysThrIleLysAs 210

```

APPLICATION NUMBER: 05/08/97/8,404B  
FILING DATE: 25-NOV-97  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/032,354  
FILING DATE: 04-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Plumer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7090  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
TELEX:  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1219 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

252 GAGGATGCTCTGCCAAGGCGGATCTAACACGCGGCTACCAATTATTA 970



```

138  sntyltyleuProValCysleuProAspProSerValIleIlephgIuThr 154
139  :::::::::: :::: :::: :::: :::: ::::
140  GCCAGTCCACACGGGTACACCTGCCCTGCTCAGAGACTTCCCCCG 451
141  155 GlyMetAsnCysTrpValThrGlyTrpGlySerProSerGluGlnAspLe 171
142  :::::::::: :::: :::: :::: :::: ::::
143  GGGATGCGCGTGGTGGTCACTGCGTGGGGCGATGTGCATGATGTAGCG 501
144  171 uleuProGluProArgGlyIleLeuGlnIysLeuAlaValProIleLeuSpt 188
145  :::: :::: :::: :::: :::: ::::
146  CTTCCACCGCCATTCTCTGAGACAGGTAAAGTCCCATATATGGAAA 551
147  188 InProLysCysAsnLeuLeuTyrSerIleSerIysAspThrGluPhgGlyTyrGln 204
148  :: :::: :::: :::: :::: ::::
149  ACCACATTGTGACGACCAAAATATACCACTTGGCGCTACAGGGAGACGAC 601
150  205 ProLysThrIleLysAsnAspMetLeuCysAlaGlyPheGluGluGly 221
151  :::: :::: :::: :::: :::: ::::
152  GTCGCGATCGTCCGAGACGACATCTGTGTCCCGG.....AACACCCG 645
153  221 slyAspAlaCysLysGlyAspSerGlyGlyProLeuValCysLeuValG 238
154  :::: :::: :::: :::: :::: ::::
155  GAGGAGACTCATCGCAGGGCGACTCCGAGGGGCCCTGTGTGTCAAGTGA 695
156  238 IyGlnSerTrpLeuGlnAlaGlyValIleSerTrpIyGluGlyCysAla 254
157  :: :::: :::: :::: :::: ::::
158  ATGCACTGCTGTGAGCGCGCTGTGTCACTGAGTGGCGCAGGCGTGTGCC 745
159  255 ArgGlnAsnArgProGlyValTyrIleArgValThrAlaHisAsnTrp 271
160  :::: :::: :::: :::: :::: ::::
161  CAGCCCAACCGGCTCGGATCATTAACCCGCTGTCACTACTGACTTGGACGTG 795
162  271 pIleHisArgIleIleProLys 278
163  :::: :::: :::: ::::
164  GATCCACCATATGTGTCCCAAA 817
165  seq_name: /cogn2_6/prodata/1/lna/5b_comb.seq:US-09-016-366A-18
166  seq_documentation_block:
167  ; Sequence 18, Application US/09016366A
168  ; Patent No. 5955431
169  ; GENERAL INFORMATION:
170  ; APPLICANT: Stevens, Richard L.
171  ; APPLICANT: Huang, Chifu
172  ; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
173  ; TITLE OF INVENTION: INHIBITORS
174  ; NUMBER OF SEQUENCES: 65
175  ; CORRESPONDENCE ADDRESS:
176  ; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
177  ; STREET: 600 Atlantic Avenue
178  ; CITY: Boston
179  ; STATE: MA
180  ; COUNTRY: U.S.A.
181  ; ZIP: 02210-4211
182  ; COMPUTER READABLE FORM:
183  ; MEDIUM TYPE: Diskette
184  ; COMPUTER: IBM Compatible
185  ; OPERATING SYSTEM: DOS
186  ; SOFTWARE: FastSeq for Windows Version 2.0
187  ; CURRENT APPLICATION DATA:
188  ; APPLICATION NUMBER: US/09/016,366A
189  ; FILING DATE: January 30, 1998
190  ; CLASSIFICATION: 530
191  ; PRIORITY APPLICATION DATA:
192  ; APPLICATION NUMBER: 60/037,090
193  ; FILING DATE: 05-FEB-1997
194  ; ATTORNEY/AGENT INFORMATION:
195  ; NAME: Plumer, Elizabeth R.
196  ; REGISTRATION NUMBER: 36,637
197  ; REFERENCE/DOCKET NUMBER: B0801/7093
198  ; TELECOMMUNICATION INFORMATION:
199  ; TELEPHONE: 617-720-3500
200  ; TELEFAX: 617-720-2441

```

```

alignment_block:
US-10-041-006-7 x US-08-978-404B-13 ..
Align seg 1/1 to: US-08-978-404B-13 from: 1 to: 1137

```

```
seq_name: /cgrn2_6/prodata/1/1na/5B_COMB.seq:US-08-978-4404B-1
seq_documentation_block:
; Sequence 1, Application US/08978404B
; Patent No. 5968782
GENERAL INFORMATION:
```

[illegible]



[illegible]

```

US-10-041-006-7 x US-08-978-404B-17 ..
Align seg 1/1 to: US-08-978-404B-17 from: 1 to: 1081

17 SerGlnArgAlaIAlaAlaThrAlaCysGlyArgProArGmetLeuAs 33
14 GGAAGCCAGCGCTACGCGGCCCTGCGCCAGGAGGCCCTGCGAGCAGT 63
33 nArGmetValGlyGlyGlnAspThrGlnGlyGlyGlyTyrProThrGln 50
64 GGGCATCTGTTGGGGGTCAAGAGGCCGCCAGGAGCAAGTGGCCCTGGCAG 113
50 aIseTlleGln.....ArgasnGlySerHisAspGlyGlySer 63
114 TGAGCGCTGAGAGTCCGCGACCGATACTGGATGCACTTCTGCGGGGCTCC 163
64 IeuTlleAlaGlnGlnTyrValLeuThrAlaAlaHisCysPhe..ArgAs 79
164 CTCATTCACCCCACTGGGTGTCTACCGCACGACATGCGTGGGACCGGA 213
79 nThrSerGlnThrSerTleuTyrGlnValLeuLeuGlyAlaArgGlnLeu 96
214 GGTCAAGAGTCTGGCGGCCCTCAGAGGTGCACCTCGGAGAGACACTCT 263
96 aGlnProGlyProHisAlaMetTyrAlaArgValArgGlnValGlnSer 112
264 ACTAC.....CAGAGCCAGCTGCTCCGCGACAGCATGATCATGTG 304
113 AsnProLeuTyrGlnGlyThrAlaSerSerHisAspValAlaLeuValG 129
305 CACCACCACTTCTACACCGGCCACAGTGGAGGAGACATCGCCCTGCTGGA 354
129 uLeuGlnAlaProValProPheThrAsnTyrIleLeuProValCysLeu 146
355 GCTGAGAGGAGCGCGTGAAGAGTCTCAGCACAGTCCACACGCTACCTGC 404
146 roAspProSerValIlePheGlnThrGlyMetAsnTyrValThrGly 162
405 CCCCTGCTCCAGACACTTCCCCCGGGGAGTCCGTGTGGGTACTGGC 454
163 TrpGlySerProSerGlnGlyAspLeuLeuProGlnProAlaIleLeuG 179
455 TGGGGGAGTGGACAATGATGAGAGCGCTCCACCGCATTTCTCTGAA 504
179 nLysIleuAlaValProIleIleAspThrProLysCysAsnLeuLys 196
505 GGAGGTGAAGGCCCCAATATGGAAACACATTGTGTGCCCAAAATACC 554
196 eLysAspThrGlnPheGlyTyrGlnProLysThrIleLysAsnAspMet 212
555 ACCTTGGCGCCCTACACAGGAGACACACGTCGCCATCGTCCGAGACATG 604
213 LeuCysAlaGlyPheGlnGlyGlyLysLysAspAlaCysLysGlyAspS 229
605 CTGTGTGCCGGG.....AACACCCGGAGGAGATCATATGCCAGGCCGATC 648
229 rGlyGlyProLeuValLysIleuValGlyGlnSerTyrLeuGlnIleGly 246
649 CGAAGGGCCCCGTGGTGTCAAGGTAAATAGGCACTGCTGAGGCGGGCG 698
246 aLleSerTyrPylGlyGlyCysAlaArgGlnAsnArgProGlyValTyr 262
699 TGGTACGTGGGGGAGAGGCGTGTGCCAGCCAAACGGCGCTGACATCTAC 748
263 IleArgValThrAlaHisHisAsnThrIleHisArgIleIleProLys 278
749 ACCCGTGTACCTACTACTGTGACATCGATCCACACATATGTCCCAAA 796
seq_name: /cgn2_6/ptodata//ina/bh_COMB.seq:US-09-008-271A-15
seq_documentation_block:
; Sequence 15, Application US/09008271A
; Patent No. 6203379

```





```

STATE: WISCONSIN
COUNTRY: U.S.A.
ZIP: 53717-1914
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patenlin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/079,970A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Leone, Joseph T.
REGISTRATION NUMBER: 37,170
REFERENCE/DOCKET NUMBER: 34506.073
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 831-2100
TELEFAX: (608) 831-2106
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 771 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 7..753
FEATURE:
NAME/KEY: misc.signal
LOCATION: 7..18
US-09-079-970A-4

```

```

alignment_scores:
Quality: 573.50      Length: 249
Ratio: 3.296         Gaps: 4
Percent Similarity: 69.880   Percent Identity: 44.578

```

alignment\_block:  
us-10-041-006-7 x us-09-079-970A-4

seg 1/1 to: us-09-079-970A-4 from: 1 to: 771

```

34 ArgMetValGlyGlyGlnAspThrGlnGluGlyGlnTrpProTrpGlnVal 50
16 AGAATCGTCGGGGGTGAGAGAGCCCGCAGAGCAAGTGGCCCTGGCAGGT 65
50 lSerIleGlnArgAsnGly.....SerHisPheCysGlyGlySerL 64
66 GAGCTTAGAGTCCAGCGCCCAATGATGATGACTTCTGGCGGGGCTCC 115
64 euIleAlaGluGlnTrpValLeuThrAlaIleHisCysPhe...ArgAsn 79
116 TCATCCACCCCGCAGTGGGTGCTGACCCAGCAGCACTCCGTGGAGACCG 165
80 ThrSerGluThrSerLeuTrpGlnValLeuLeuGlyAlaArgGlnLeuVal 96
166 GTCAGAGATGTCGGCCCTCAGGTCGACACTGCGGGAGCAGACACTCTA 215
96 lGlnProGlyProHisAlaMetTrpAlaArgValAlaGlnValGlnSer 113
216 CTAC.....CAGGACGAGTCTGCTCCGCTCAGCAGGATCATCTGTC 256
113 snProLeuTrpGlnGlyThrAlaSerSerAlaAspValAlaLeuValGln 129
257 ACCCAGAGTCTACACCGCCCGCAGATGCGAGCGAGCATCGCCCTGCTG 306

```

```

130 LeuGluAlaProValProPheThrAsnTrpIleLeuProValCysLeuPr 146
307 CTGGAGAGCCCGGTGAAGTCTTCACGACGTCACAGCGGTACCTGGCC 356
146 oaSPProSerValIlePheGlnTrpGlyMetAsnCysTrpValThrGlyT 163
357 CCTGCTCAGACACTTCCCGGGGATGCCGCTGCTGGTACACTGGCT 406
163 rplYleSerProSerGluGlnAspLeuLeuProGluProArgIleLeuGln 179
407 GGGCGCATGTGACAAATGATGAGCGCTCCACCGCCATTCTCTGAGG 456
180 LysLeuAlaValProIleIleAspThrProLysCysAsnLeuLeuTyrSe 196
457 CAGTGAAGTCCCGCATATGGAAGAAACCATTTGTGAGCAAAATACCA 506
196 rlyAspThrGluPheGlyTyrGlnProLysThrIleLysAsnAspMetL 213
507 CCTGGCGCTACACGCGAGACAGCGTCCGATCGTCCGTGACGACATGC 556
213 euCysAlaGlyPheGlnGluGlyLysLysAspAlaCysLysGlyAspSer 229
557 TGTGTCCCGGG....AACCCCGGAGGAGTCAATGATGCGAGGCGACTTC 600
230 GlyGlyProLeuValCysLeuValGlyGlnSerTrpLeuGlnAlaGlyVa 246
601 GAGGGCCCCCTGTGTGCAAGTGAATGCGCACTGGCTCAGAGCGGCGCT 650
246 lIleSerTrpGlyGlyCysAlaArgGlnAsnArgProGlyValTyrI 263
651 GGTACACTGGGGGAGGAGGCTGTGCCAGCCCAACGCGGCTGCGCATCTCA 700
263 LeuValThrAlaHisHisAsnTrpIleHisArgIleLeuProLys 278
701 CCCGTCACTACTACTGTGACTGGATCCACCATGATGTCCTCCCAA 747
seq_name: /cgn2-6/ptodata/1/lna/5B.COMB.seq-us-08-978-404B-4

```

seq documentation block:

Sequence 4, Application US/08978404B  
Patent No. 5968782

GENERAL INFORMATION:  
APPLICANT: Stevens, Richard L.  
TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
NUMBER OF SEQUENCES: 74  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA

COUNTRY: U.S.A.  
ZIP: 02210-2211

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/978,404B  
FILING DATE: 25-NOV-97  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/032,354  
FILING DATE: 04-DEC-1996  
ATTORNEY/AGENT INFORMATION:

NAME: Plumer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7090  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441

661 A.....AATACCGAGGAGCACTCCTGCCAGGAGACTTCAGGGGGCCACC 704  
233 euValcysLeuValIglYnserTrrPleuGlnalagIValIIeserTrp 249  
||||| ||| :|||:  
705 TGCTGTGCAAAGAAGGAGGCTACTGGCTGCACAAGCAGAGAGTGTCACGCTGG 754  
250 gIaGlIGlICVaslaIaaArgInImSnrrProGIlvaITrrIIeaArwaIrh 266

755 GGTGAGGGGCTGCSCAAGAGCCCAATCGTCTCGGACATTTCACACC GGGTGAC 804  
| :::::::::::::::::::::  
266 TATAATSHASATPRTPIlenIleArgIleProLysLeuSlnPhcGlnr 283  
| :::::::::::::::::::::  
805 GTACTACSTGGATGGATTACACCGTATGCTCCTCAGGGTTCTCTGAAC 854

283 rolaargleu 286  
||:::||:::

[illegible]

```

seq_documentation_block:
  Sequence ID: Application US/09016366A
  Patent No. 5955431
  GENERAL INFORMATION:
    APPLICANT: Stevens, Richard L.
    APPLICANT: Huang, Chifu
    TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
    NUMBER OF INVENTION: INHIBITORS
    NUMBER OF SEQUENCES: 65
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
      STREET: 600 Atlantic Avenue
      CITY: Boston
      STATE: MA
      COUNTRY: U.S.A.
      ZIP: 02210-2211
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Diskette
    COMPUTER: IBM Compatible
    OPERATING SYSTEM: DOS
    SOFTWARE: FASTEO for Windows Version 2.0
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/016.366A
    FILING DATE: January 30, 1998
    CLASSIFICATION: 530
    Prior APPLICATION DATA:
      APPLICATION NUMBER: 60/037,090
      FILING DATE: 05-FEB-1997
    ATTORNEY/AGENT INFORMATION:
      NAME: Plumer, Elizabeth R.
      REGISTRATION NUMBER: 36,637
      REFERENCE/DOCKET NUMBER: B0801/7093
      TELECOMMUNICATION INFORMATION:
        TELEPHONE: 617-720-3500
        TELEFAX: 617-720-2441
      TELEX:
    INFORMATION FOR SEQ ID NO: 16:
      SEQUENCE CHARACTERISTICS:
        LENGTH: 1154 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: single
        TOPOLOGY: linear
        MOLECULE TYPE: cdna
  US-09-016-366A-16

alignment_scores:
  Quality: 572.50      Length: 274
  Ratio: 3.078         Gaps: 4
  Percent Similarity: 67.883      Percent Identity: 43.796

alignment_block:
  US-10-041-006-7 x US-09-016-366A-16

```





ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	1 (sites)
JOURNAL MEDLINE	Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)
PUBMED	99279636
REFERENCE	10349633
AUTHORS	2 (sites) Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL MEDLINE	20499374
PUBMED	11042159
REFERENCE	3 (sites) Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsumi,T., Tashiro H., Itoh M., Sumi.N., Ishii.Y., Nakamura.S., Hazama.M., Nishine.T., Harada.A., Yamanoto.R., Matsumoto.H., Sakaguchi.S., Ikegami.T., Kashiwagi.K., Fujiwaka.Y., Inoue,K., Togawa.Y., Izawa,M., Ohata,E., Watabiki.M., Toneida.Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL MEDLINE	20530913
PUBMED	11076861
REFERENCE	4 (sites) The RIKEN genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 665-690 (2001)
JOURNAL MEDLINE	5 (phases I to 1323)
PUBMED	Nature 409, 665-690 (2001)
REFERENCE	5 (phases I to 1323)
AUTHORS	Akachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Butt,C., Carninci,P., Fukuda,S., Fukushima,Y., Furuno,M., Hanagaki,T., Hara,A., Hayasaka,N., Hill,D., Hiramoto,K., Hiroaka,T., Horii,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Koijima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numata,K., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shingawara,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
JOURNAL	Submitted
TITLE	Submitted (10-JUL-2000) yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGATCCACAGACTTCCTTTTCTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGATTCCTGCATTAAATAAATTAATCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size approximately 0.5 and 3 kb



Ratio: 5.398 Gaps: 0  
Percent Similarity: 96.241 Percent Identity: 95.489

alignment block:

US-10-041-006-7 x AM170323/rev ..

Align seg 1/1 to reverse of: AM170323 from: 1 to: 602

```

158 CysTrpValThrGlyTrpGlySerProSerGluGluAspLeuLeuProGly 174
|||||
602 TGCTGGGTCTACGCTGCGCGCACCTAGTGAAGTAAAGACCTCTGCGCGA 553
174 uProArgIleLeuGlnLysLeuAlaValProIleIleAspThrProLysC 191
|||||
552 ACAGCGGATCTCGACAGTACCTGCTGCGCCATCATCGAACACCCAAAG 503
191 yAsnLeuLeuTyrrSerLysAspThrGluPheGlyTyrrGlnProLysThr 207
|||||
502 GTACACATGCTCTACACCAAGACACCGAGTTGGCTTACCAACCCAAACC 453
208 IleLysAsnAspMetLeuLeuGlnGlyPheGlnGluGlyLysLysAspAl 224
|||||
452 ATCAAGATGACATGCTGTCGCGCGCTTCGAGAGGAGGCAAGAGATGC 403
224 acYsLysGlyAspSerGlyGlyProLeuValCysLeuValGlyGlnSerT 241
|||||
402 TTGCAAGGGGAGTACGCGCGCGCCCTGGTGTCCTCGTGGGTCACTGCT 353
241 rPheGlnIleGlyValIleSerTrpGlyGluGlyCysAlaArgGlnAsn 257
|||||
352 GGGTGGAGGGGGGGGTGATCAGTGGGGTGAAGGGCTGTCGCCCGCAAGC 303
258 ArgProGlyValTyrrIleArgValThrAlaHisHisAsnTrpIleHisAr 274
|||||
302 CGCCGAGGTGATACATCCGTGCACCGCGCCACCACTGATCATGTCG 253
274 gIleIleProLysLeuGlnPheGlnProAlaArgLeuGlyGlnLys 290
|||||
252 GATCATCCCAAACTGCTGACGACGAGGAGTGGCGCGCGCGCAAG 204
seq_name: gb_est1:AM084167

```

seq documentation block:

LOCUS AM084167 601 bp mRNA linear EST 14-OCT-1999  
DEFINITION xc48c07.x1 NCI\_CGAP\_Eso2 Homo sapiens CDNA clone IMAGE:2587500 3'  
similar to TR:043342 O43342 SP0011A ;, mRNA sequence.

ACCESSION AM084167 GI:6039319

VERSION

WORDS

ORIGIN

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 601)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: rgaps@femail.nih.gov  
Tissue Procurement: Nan Hu, M.D., Ph.D., Mark Roth, M.D., Phillip  
Taylor, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www-bio.llnl.gov/bdrp/image/image.html

FEATURES

Possible reversed clone: polyT not found  
Seq primer: -40UP from gibco  
High quality sequence stop: 408.  
Location/Qualifiers

source

```

1. 601
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2587500"
/clone_lib="NCI_CGAP_Eso2"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B"
/note="Organ: esophagus; Vector: pcMV-SpOrf6; Site:1: Salt
; Site:2: NCI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.1 Kb. Life Technologies catalog
#: 11502-010"

```

BASE COUNT 114 a 157 c 203 g 127 t

ORIGIN

alignment\_scores:

Quality: 667.00 Length: 137  
Ratio: 5.131 Gaps: 3  
Percent Similarity: 94.891 Percent Identity: 93.431

alignment\_block:

US-10-041-006-7 x AM084167/rev ..

Align seg 1/1 to reverse of: AM084167 from: 1 to: 601

```

157 AsnCysTrpValThrGlyTrpGlySerProSerGluGluAspLeuLeuP 173
|||||
601 AACTGCTGTTCTACGCTGCGCGCGCTTCGAGAGGAGGCAAGAG 552
173 rGluProArgIleLeuGlnLysLeuAlaValProIleIleAspThrPro 189
|||||
551 CGGAACCGCGGATCTCGAGAACTCGCTGCGCCATCATCGAACACTC 502
190 Lys CysAsnLeuLeuTyrrSerLysAspThrGluPheGlyTyrrGlnPro 206
|||||
501 AAGGTGCAACCTGCTTACACCAAGACACCGAGTTGGTGTACCAACCA 452
206 yStrIleLysAsnAspMetLeuLeuGlnGlyPheGlnGluGlyLysLys 222
|||||
451 AAACCATCAAGATGACATGCTGTCGCGCGCTTCGAGAGGAGGCAAGAG 402
223 AspAlaCysLysGlyAspSerGlyGlyProLeuValCysLeuValGlyG 239
|||||
401 GATCCCTGCAAGGGGAGTACGCGCGCGCCCTGGTGCTCGTGGGTCA 352
239 nSerTrpLeuGlnIleGlyValIleSerTrpGlyGluGlyCysAlaArg 256
|||||
351 GTCGTGCTGCAAGGGGAGTACGCTGCGCGCGCTTCGAGAGGAGGCTGCC 302
256 lnaAsnArgProGlyValTyrrIleArgValThrAlaHisHisAsnTrpI 272
|||||
301 AGAACCGCCGAGGTGCTACATCCGTGCACCGCGCCACCACTGATCAT 252
273 HisArgIleIleProLysLeuGlnPheGlnProAlaArgLeuGlyGly 289
|||||
251 CATGATCATCCCAAACTGCAAGTCAACCAAGAGGAGTGGCGCGCC 202
289 lnyLys 290
|||||
201 AGAAG 197
seq_name: gb_est1:BB115142

```

seq documentation block:

LOCUS BB115142 625 bp mRNA linear EST 18-OCT-2001  
DEFINITION BB115142 RIKEN full-length enriched, adult male urinary bladder Mus  
musculus CDNA clone 9530049D14 3', mRNA sequence.

ACCESSION BB115142 GI:16261796

VERSION BB115142.2

KEYWORDS EST

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;





Email: c9apbs-r@mail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Inocyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LNL1243 row: 1 column: 14  
 High quality sequence stop: 704.  
 Location/Qualifiers

FEATURES  
 source  
 1. 889  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5101213"  
 /clone\_lib="NCI\_CGAP\_L19"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: Liver; Vector: PCMV-SPORT6; Site.1: NotI;  
 Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.9 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library."  
 COUNT 188 a 249 c 242 g 210 t  
 GCIN

alignment\_scores:  
 Quality: 594.50 Length: 254  
 Ratio: 3.231 Gaps: 9  
 Percent Similarity: 72.441 Percent Identity: 50.394

alignment\_block:  
 US-10-041-006-7 x B1218460 ..

Align seg 1/1 to: B1218460 from: 1 to: 889

```

24 ThrAlaCysGlyArgProArgMetLeuAsnArgMetValGlyGlyGlnAs 40
   |||||
7 TCAGGTGTGGAGCGTCCAGCATGACCATGCGTCCAGGACAGAGA 56
   |||||
40 PThcGlnGluGlyLutPProTgPglValSerTleGlnArgGnglys 57
   |||||
57 TGCCAGCTGGGCGGCTGGCTGGAGGTAGCGTGGGAGATGGGG 106
   |||||
57 eRhISpHeCysGlyLysSerLeuIleAglGlnTgPglValLeuThrAla 73
   |||||
107 GCAGAGTGTGGGAGATCTGTGATGCTGAGAGCTGGGTGACGTCT 156
   |||||
74 AlaHIScysPheArgAsnThrSerGluThrSerLeuTyrGlnValLeu 90
   |||||
157 GCCATTGCTTCAACCAAGGCCAGTCCCTGTCTATCAACAGTACGTCT 206
   |||||
90 uGlyAlaArgGlnLeuValGlnProGly.....ProHISAlaMetTyrA 105
   |||||
207 GGGTACCATCTCTCTACCTGAGGACATGACCCAGAGAGCTGGAGAG 256
   |||||
105 laaTgvalAArgGlnValGlnSerAsnProLeuTyrGlnGlyThrAla... 120
   |||||
257 CT...GTGGCCAGCTTATCAAGCACCAGCTATTCAGCGAGAGACAC 303
   |||||
121 SerSerAlaAspValAlaLeuValGlnLeuGlnAlaProValProPhe 137
   |||||
304 AGCATGGAGACATTTGCCCTGTGACCTGCTTACCCATCTCCCTCAA 353
   |||||
137 rAsHtyrIleLeuProValCysLeuProAspProSerValIlePheGlu 154
   |||||
354 TGAATACATGCTTCCAGTCTCCCTCCGAACCTGGGAGACCCCTGATC 403
   |||||
154 hrGlyMetAsnCysTgPglValThrGlyTgPglSerProSerGluAsp 170
   |||||
404 CTGGCACCAGTGTCTGGGTGATCTGGGAGACACATTTGGCACAAATCA 453
   |||||
171 LeuLeuProGluProArgIleLeuGlnLysLeuAlaValProIleIleAs 187

```

```

|||||
454 CCCTCCACACACCTTTACCTGCGAGAGATTGCAAGTGTCTCTATTGA 503
   |||||
187 pThrProLysCysAsnLeuLeuTyrSerTyrAspThrGlnuphGlyTyrG 204
   |||||
504 TGCCGAGACCTGCAATATCACTACTACAGAGAGACTCATCTCCCTGGACGG 553
   |||||
204 InProLysThrIleLysAsnAspMetLeuCysAlaGlyPheGluGluGly 220
   |||||
554 AGCA...GTATCTCTTGAAGGCATGCTGTGTCTGTTCCAGGAAGGC 600
   |||||
221 LysLysAspAlaCysLys...GlyAspSerGlyGlyProLeuValCysLeu 237
   |||||
601 AAGAGAGATCTGTGAATGCTGATCCGAGAGTCCCTAGTCTGTGACA 650
   |||||
237 aLgIlyGlnSer...TgPglValIleSerTgPglGlu...G 252
   |||||
651 TTCATGACTGTCTGATCCAGGAGGAGGTGCTGATGGGATCTGTGAT 700
   |||||
252 LysAlaArgGlnAsnArgPro.....GlyVal 261
   |||||
701 TGTGATCTGTTCAGAGAGCCAGCTGTTCACCCAGCAAGCTCATGTCT 750
   |||||
262 TyrIle 263
   |||||
751 TACATT 756

```

seq\_name: gb\_est2:BG962187

seq\_documentation\_block:

LOCUS BG962187 802 bp mRNA linear EST 12-JUN-2001  
 DEFINITION 602826954F1 NCI\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4981789 5',  
 mRNA sequence.  
 ACCESSION BG962187  
 VERSION BG962187.1 GI:14349824

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Comment: Email: c9apbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Inocyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.llnl.gov>

Plate: LNL10983 row: 1 column: 14

High quality sequence stop: 675.

Location/Qualifiers

1. 802

/organism="Mus musculus"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:4981789"

/clone\_lib="NCI\_CGAP\_Co24"

/lab\_host="DH10B (T1 phage-resistant)"

/note="Organ: colon; Vector: PCMV-SPORT6; Site.1: NotI;  
 Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.6 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT

ORIGIN

179 a 213 c 234 g 176 t

alignment\_scores:

Quality: 591.50 Length: 248

Ratio: 3.250 Gaps: 6  
Percent Similarity: 73.387 Percent Identity: 49.597

Alignment\_block:  
US-10-041-006-7 x BG962187 ..

Align seg 1/1 to: BG962187 from: 1 to: 802

```

24 ThrAlaGlyArgProArgMetLeuAsnArgMetValGlyGlnAsp 40
29 TGAAGTGTGGAGCTCCAGCATCAGCAGATCGTGTGAGCAGACA 78
40 PThrGlnGlyGluTrpProTrpGlnValSerIleGlnArgAsnGly 57
79 TCCCGACGCGGCGGTGGCTTGGCAGGTCCGCTGGCGGAGATGGGG 128
57 eRh1sPheCysGlySerLeuIleAlaGlnGlnTrpValLeuThra 73
129 CCGACGTGTGGGAGATCTGATTCGTGAGAGCTGGGTGTGACTGCT 178
74 AlaHisCysPheArgAsnThrSerGluThrSerLeuTyrlValLeu 90
179 GCCCATTCCTTCAACAGGCGCAGTCCCTGTATCTACACGTACTGCT 228
90 uGlyAlaArgGlnLeuValGlnProGly.....ProHisAlaMetTyra 105
229 GGGTACCATCTCCTCTACCTCCAGGACAAATGAGCGAAGAGAGAG 278
105 IaArgValArgGlnValGlnSerAsnProLeuTyrlGlnGlyThra 120
279 CT...GTGGCCAGATTATCAAGCACCCAAAGCTATTCACGCGAGC 325
121 SerSerAlaPheValAlaLeuValGlnLeuGlnAlaProValProPhe 137
326 AGCAGTGGAGACATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 375
137 rAsnTyrlLeuProValAlaCysLeuProAspProSerValIlePhe 154
376 TGACATCATGCTTCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 425
154 hGlyMetAsnGlyTrpValIleGlyTrpGlySerProSerGlnGlu 170
426 CTGGCACCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 475
171 LeuLeuProGlnProArgIleLeuGlnValAlaValProIleLeu 187
476 CCGCTCCCAACAACCTTACCTGCGAGAGAGTGCAGGTCTCTCTTA 525
187 PThrProLysCysAsnLeuLeuTyrlSerLysAspThrGlnPheGly 204
526 TGGCGAGACCTGCATTAATACTACAGAGAACTCCATTCTGCGACG 575
204 InProLysThrIleLysAsnAspMetLeuCysAlaGlyPheGlnGly 220
576 AGCA...GTCACTCTTGAAGCATCTGTGCTGCTGCTGCTGCTGCT 622
221 Lys-LysAspAlaCysLys-GlyAspSerGlyLysProLeuValCys 236
623 AAGACAGAGATCTTGAATAGTGATTCGCGAGAGTCCCTAGCTGTGC 672
237 ValGlyLeuSerTrpLeuGlnAlaGlyValIleSerTrpGlyGln 253
673 ATAAATGATGTGTGATTCAGGAGGAGAGTGGGAGATCTGATTG 722
253 AlaArgGlnAsnArgProGlyValTyrlLeuArgValThr 266
723 TGCTCTGTACACAGCGGCTGTATAACACAGCATGTACGT 762
seq_name: gb_est1:AL551470

```

seq\_documentation\_block:  
LOCUS AL551470 999 bp mRNA linear EST 16-FEB-2001  
DEFINITION AL551470 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CS0D1064YF02 5

prime, mRNA sequence.  
ACCESSION AL551470 GI:12889449  
VERSION AL551470.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 999)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr.

#### FEATURES

source  
1..999  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="LTI\_NFL006\_PL2"  
/tissue\_type="placenta"  
/note="Vector: PCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifestech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 179 a 331 c 295 g 190 t 4 others

alignment\_scores:  
Quality: 588.50 Length: 288  
Ratio: 2.987 Gaps: 9  
Percent Similarity: 68.403 Percent Identity: 42.708

alignment\_block:  
US-10-041-006-7 x AL551470 ..

Align seg 1/1 to: AL551470 from: 1 to: 999

```

6 AlaValProLeuLeu.....LeuLeuCysPheGlySerGlnArg 19
115 GCTGTGGCCATCTGCTATCTGATTAATCTCGGTGCGGAGACGAGC 164
19 GAlaLysAlaAlaThrAlaCysGly...ArgProArgMetLeuAsnArg 35
165 GGAAGGCGCAGAAAGTCCCTCGGTGGTGGCCCCCAACA....CGCA 208
35 eValAlaGlyLeuAspThrGlnGlnGlyLysTrpProTrpGlnValSer 51
209 TCACAGGTGGCAGCAGTCCAGTCCGCGGTGCTGAGTCCCTGCGAGTGC 258
52 IleGlnArgAsnGlySerHisPheCysGlySerLeuIleAlaGlnGly 68
259 ATCACTATGAAGGCGTCACTGTGTGTGCTGCTGCTGCTGCTGCTG 308
68 nTrpValLeuThrAlaAlaHisCysPheArgAsnThrSerGluThrSer 85
309 GTGGGTGCTGTACAGTCTGCTACGCTTCCCGAGCAGACCAAGAGAG 358
85 euTyrlValLeuLeuGlnAlaArgGlnLeuValGlnProGlyProHis 101
359 CCTATGATGTCMAAGCTGGGGGCCACGCTAAGTACTCTACTCCGAGGC 408
102 AlaMetTyrlAlaArgValArgGlnValGlnSerAsnProLeuTyrlGln 118

```

JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
20499374	
PUBMED	11042159
REFERENCE	
AUTHORS	3 (sites)
	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kondo,H., Akiyama,J., Nishi,K., Kitsumai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishize,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Keshiwaiki,K., Fujikawa,S., Inoue,K., Togawa,Y., Itawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.
TITLE	RIKEN integrated sequence analysis (RISA) system-364-format
JOURNAL	sequencing pipeline with 384 multipillarary sequencer
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)
PUBMED	20530913
REFERENCE	11076861
AUTHORS	4 (sites)
	The RIKEN genome Exploration Research Group Phase II Team and the FANCOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
20499374	5 (bases 1 to 1629)
PUBMED	11042159
REFERENCE	
AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,

2 (sites)  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

misc\_feature

Genome Res. 10 (10), 1617-1630 (2000)

20499374

11042159

3 (sites)

Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,Y., Nishik,K., Kitsuana,T., Tashiro,H., Itoh,M., Sumi.N., Ishii.Y., Nakamura.S., Hazama.M., Nishine.T., Harada.A., Yamamoto.R., Matsunoto.H., Sakaguchi.S., Ikegami.T., Kashiwagi.K., Fujitake,S., Inoue.K., Togawa.Y., Izawa.M., Ohara.E., Watabiki,M., Toneka.Y., Ishikawa.T., Ozawa.K., Tanaka.T., Matsura.S., Kawai.J., Okazaki.Y., Muramatsu.M., Inoue.Y., Kiru,A. and Hayashiakazi,Y.

Riken integrated sequence analysts (RISA) system-384-format sequencing pipeline with 384 multipipillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

20530913

4 (sites)

11076861

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5 (bases 1 to 1629)

Aoachi,U., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bull,C., Carninci,P., Fukuda,S., Fukushima,Y., Furuno,M., Hanagaki,T., Hara,A., Hayasu.N., Hill,D., Hiramoto,K., Hiroaka.T., Hori.F., Hume.D., Imotani,K., Ishii,Y., Itoh.M., Izawa.M., Kasukawa,T., Kato,H., Kawai,J., Kojima.Y., Konno.H., Kouda.M., Koyra.S., Kurihara.C., Matsuyama.T., Miyazaki,A., Nishi,K., Nomura,K., Numataki,R., Ono.M., Okazaki,Y., Okido,T., Owa.C., Quackenbush,J., Saito,H., Salto,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schreml,L., Shbate,K., Shibata.Y., Shinagawa,A., Shiraki,T., Soejima.Y., Suzuki.H., Tagami.M., Tagawa.A., Takahashi,F., Tanaka,T., Teijima.Y., Toya.T., Yamamura.T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino.M., Muramatsu,M. and Hayashiakazi,Y.

Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN). Laboratory for Genome Exploration Research Group, RIKEN Genomic Science Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

Encyclopedia was prepared and sequenced in Mouse Genome ENCODE library Project of Genome Exploration Research Group in Riken National Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGGACAGAAGATCCCAAGGCCTTTTTCCTTTTTTTTAA 3'], cDNA was prepared by using Trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGGACAGAATTCCTATTATTAATTAATCCCCCCCCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

1. 1629  
/clone\_11b=-RIKEN full-length enriched mouse cdna library  
/note="data source:SPR, source key:Q9P5D1, evidence:ISS  
similar to PROSTASIN PRECURSOR (EC 3.4.21.-) (CHANNEL  
ACTIVATING PROTEASE 1)"

180 GCCTGGCCCTGTGGATTCCACGATGCGAGTGAGCCCTGCTCAGCCAC 229

[illegible]

TITLE  
JOURNAL  
COMMENT

Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasakawa, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M., and Hayashizaki, Y.  
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@gsic.riken.go.jp,  
URL: <http://genome.gsic.riken.go.jp/>

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to generate full-length cDNA libraries for rapid discovery of new

genes. *Genome Res.* 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, I., Izawa, M., Uchida, E.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuuru  
Watahiki, M.

S., Kawal, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kida, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, T. (2000)

Computer-based methods for the mouse full-length cDNA  
, Y. and Hayashizaki, Y.

Computer-Aided Molecular Design for the Design of a  
Encyclopedia: real-time sequence clustering for construction of a  
redundant cDNA library Genome Res 11 (2): 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaoka, I., Aizawa

K., Fukuda, S., Hara, A., Itoh, M., Kawada, J., Shibata, A. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome*. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

location/qualifiers  
e mouse tissues.

Location/Year: 1962  
1. 626  
Accession="Mus musculus"

```

/organisms= mus musculus
/db_xref="taxon:10090"
/ncbi_refseq="NC020402.1"

```

```
/clone_lib="RIKEN full-length enriched, adult male urinar
```

bladder"/sex="male"

```
/tissue_type="urinary bladder"  
/dev_stage="adult"
```

```
/lab_host="DH10B"  
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
```

prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken

Project of genome exploration research, the National Genomic Sciences Center and Genome Science Laboratory in Nivn Division of Environmental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5'-GAGAGAGAGATCCAGAAGCTCTTTTTTTTTTTVN 3'], cDNA was

prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length

cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 370.0. Second

strand cDNA was prepared with the primer adapter of sequence 5' GAGAGAGAGATTCGAGTTAAATTAATCCCCCCCCC

3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from

Lambda FLC I. "

140 a 187 c 177 g 122 c

scores:









ACCESSION	AK004939
VERSION	AK004939.1 GI:12836502
KEYWORDS	HTC; GAP trapper.
SOURCE	Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA, clone_1lb:RKEN full-length enriched mouse cDNA library clone:1300008a22.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (sites)
AUTHORS	Carninci,P. and Hayashizaki,Y.
JOURNAL	High-efficiency full-length cDNA cloning
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)
PUBMED	99279253
REFERENCE	10349636
AUTHORS	2 (sites) Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL MEDLINE PUBMED	Genome Res. 10 (10), 1617-1630 (2000)
REFERENCE AUTHORS	Shibata,K., Itoh,M., Alzawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsumai,T., Tashiro,H., Itoh,A., Sumi,N., Ishii,Y., Nakamura,S., Hazama,N., Nishino,T., Harada,S., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiyagi,K., Fujiwara,K., Inoue,K., Togawa,I., Izawa,M., Ohtate,E., Watanabe,M., Tameda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuyura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,T., Kita,A. and Hayashizaki,Y.
TITLE	Riken integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multipillarary sequencer
JOURNAL MEDLINE PUBMED	Genome Res. 10 (11), 1757-1771 (2000)
REFERENCE AUTHORS	20530913 11076861
TITLE	4 (sites) The RIKEN genome Exploration Research Group Phase II Team and the FANTOM Consortium.
JOURNAL MEDLINE PUBMED	Functional annotation of a full-length mouse cDNA collection
REFERENCE AUTHORS	Nature 409, 665-669 (2001) 5 (pages 1 to 3030)
TITLE	Adachi,J., Alzawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Butt,C., Carninci,P., Fukuda,S., Fukushima,Y., Furuno,M., Hanagaki,T., Hara,N., Hayasun,N., Hill,D., Hiramoto,K., Hirotsu,T., Horii,F., Hume,D., Imochi,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sobabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamanura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
JOURNAL MEDLINE PUBMED	Direct Submission Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGGAGAAGCGCCGCCCAACATCGAGTGTCTTTTCTTTTTCVN 3'] cDNA was prepared by using reverse transcriptase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGGAGAAGCGCTCACAAGCTCATTAATTATTTAACCCTCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.
FEATURES	location/Qualifiers
source	1..3030 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="MGI:MGI:1904397" /db_xref="taxon:10090" /clone="1300008A22" /sex="male" /tissue_type="liver" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult"
CDS	91..2490 /note="data source:SPRR, source key:095519, evidence:TSS"

homolog to DJ1170K4.4 (NOVEL PROTEIN) (FRAGMENT)

putative" /codon\_start=1

/protein\_id="BAB23684.1"

/db\_xref="GI:12836503"

/translation="MPTTEVPOAADGGADGDEAEPEGRKPRKNTKRNRDYVR

FTPLIVLALVAGVLMFLGKAEVTSQYSGSLRVLNHFSSODLGRRESIAER

SESAKQKMLDELVASTRIGTYNSSSVSPGGSGPLTGFPLTDIPEORLTLSPDV

VRELDELNSNSTLASITYKTEYEVDEGVIIEASNDIVINSLTGCTRIYVPG

QVPLKGPDDQTTSCMLHLLQPEDLMKIVLEWTRVDCRDMATDAGFLERLITS

VYGSRQEPVMEVLAAGSVAVVWKGMHSYDPEFLLSVSAFODQVNLLEGRID

TOGRLRTPYSPYSPESTHCSMLTVPISLDYGLALFMDALRQKYNRLCTGOMMI

ONRLCGFRTLOPAERIIMVASDGVITNFSOISLTPGVQYYSLYNDSDCPGEF

LCSYNGLCVPACDGIKDCPNGLDERNCVCRAMQSCDSDSTCISLPRYCDROPCLNGS

DEEGCCDGVPCGPTTPOCEDRSCYKRNPECDQSDSDSDHDCGIGLSSRIY

GGTVSSBGEHPWQASLQIRGRHICGGLALNDRWITDAHCFQDSMASPFLMTVFLTK

MKONSRLMPGEVSEFVSRFLFLPHYHEEDSDHYDALQLDHPVYSAIVRVCCLPASH

FFEPGQCMWTGMAQREGPVNSNTLQKVDVQLPDCLSEARLYQVSPMLCAGYRK

GKDKACGGDGGGPLYCREPSGRWFLAGLVSMGICGRPNFEGYTRVTRVIMIQVL

T"

polya\_signal 3007 . 3012

polya\_site /note="putative"

BASE COUNT 609 a 903 c 883 g 635 t

ORIGIN

alignment\_scores: Quality: 555.00 Length: 254

Ratio: 3.227 Gaps: 5

Percent Similarity: 67.717 Percent Identity: 43.701

alignment\_block:

US-10-041-006-7 x AK004939 ..

Align seg 1/1 to: AK004939 from: 1 to: 3030

26 CysGLAArgProArgMetLeuAsnArgMetValGlyGlnAspThrG1 42

1756 TGTGGCCTCCAGGGCCTCTCCAGCCGATGTGGGGGAGCCGTGCTC 1805

42 ngluglylurProTPrGlnValSerIleGlnArgAsnGlySerH1sp 59

1806 CGAGGGGTGAGTGGCCATGGCAGCCAGCCTCCAGATCGGGGTCCAGACA 1855

59 heCysGlySerIleValGlnGlnTPrValLeuThraAlaHis 75

1856 TCTGTGGGGGCTCTCATCGCTGACCGCTGGGTATACGGCCGCCAC 1905

76 CysPheArgAsnThrSerGlnThrSer.....LeuTyrGlnValLeu 90

1906 TGGTTCAGAGGAGACAGATGGCTCCCGAAGCTGTGAGCCGTGTCT 1955

90 uGlyAlaArgGlnLeuValGlnProGlyProHisAlaMetTyrAlaArgV 107

1956 GGGAAAGATGCGGACAACTCGCGGTGGCAGGAGGTGCTTCAAGG 2005

107 a1ArgGlnValGlnSerAsnProLeuTyrGlnGlyThraLaserSerAla 123

2006 TGAAGCGTGTCTCTGACACCGCTACACAGAGGAGACGACATGACTAC 2055

124 AspValAlaLeuValGlnLeuGlnAlaProValProPheThrAsnTyr11 140

2056 GACGTGGCCCTGCTGACGCTCCAGCACCCTGTGGTACTCGGCCACTGT 2105

140 eLeuProValCysLeuProAspProSerVal11lePheGlnThrGlyMetA 157

2106 GCGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2155

157 snCysTrpValThrGlyTPrGlySerProSerGlnGlnAspLeuPro 173

2156 ACTGCTGATCACAAGGCTGGGAGCCACGAGAGGGTGT.....CCG 2199

174 GluProArgIleLeuGlnLysLeuAlaValProIleIleAspThrProly 190

2200 GTGAGCAACACCTGTGACAGAGGTGAGCTGACGTGGTCCCTCAGACCT 2249

190 scYasAsnLeuLeuTyrSerIleCysAspThrGluPheGlyTyrGlnProlyst 207

2250 CTGCAGTCAGGCCCTACCCG.....TACCAG..... 2274

207 hTleLysAsnAspMetLeuCysAlaGlyPheGlnGlnGlyLysLysAsp 223

2275 ..GTGTCCCAAGCATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2322

224 AlaCysLysGlyAspSerGlyGlyProLeuValCysLeuValGlyGlnse 240

2323 GCGTCCCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2372

240 r...TPrLeuGlnAlaGlyValIleSerTPrGlyGlyGlyCysAlaArg 256

2373 CCGGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2422

256 InAsnArgProGlyValTyrTleArgValThraAlaHisAsnTPrIle 272

2423 CCAATTCCTTGTGGGCTCTACACCCGCTGTACACGCTGTATCAACTGATC 2472

273 HisArgIleIle 276

2473 CAGCAGTGTCTG 2484

Wed Aug 14 08:39:18 2002

us-10-041-006-7.p2n.rst

Page 16